Gigahertz Bandwidth and Nanosecond Timescales: New Frontiers in Radio Astronomy Through Peak Performance Signal Processing

Thesis by
Ryan Mckay Monroe

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 7, 2018

Ryan Mckay Monroe
ORCID: 0000-0002-0026-4546

ACKNOWLEDGEMENTS

I'll never forget the moment that I decided to embark on my PhD. I was driving up to Owens Valley Radio Observatory with my (at the time) class professor, Gregg Hallinan. Talking about his challenges with the ionosphere, we tried to calculate how much power it would take to shape the ionosphere into a giant low-frequency radio lens for our array[1]. It was research-love at first sight and began five years of mad (radio) science. With a simply staggering breadth and depth of expertise, Gregg is one of the most brilliant individuals I have met. His can-do attitude resulted in many hours standing in front of the white-board inventing new ways to measure the ionosphere and detect satellites. As an advisor, the tremendous support he gives all of his students in all of their endeavors is invaluable. I would not have started my PhD if not for him.

The CASPER community is a tremendous resource to radio astronomers around the world, and I'm no exception. The list of people is too long to name, but a special mention to Dan Werthimer, Jack Hickish, and Glenn Jones. I am very lucky to have met Jonathon Kocz, and even luckier when he started researching at Caltech as well–his ideas have bailed me out of many an FPGA problem.

The Xilinx Inc. University Program has generously donated every FPGA used in this thesis, which was crucial for all of our instruments.

The excellent research staff at Owen's Valley Radio Observatory carry the weight of several instruments and have been vital in problem-solving, infrastructure, support, trench-digging, cable-running and the thousand other vital tasks needed to make a working radio observatory. I am very grateful for their expert advice.

I appreciate our research group – in addition to being the father of modern radio-astronomy, Sandy Weinreb provided insights greatly simplifying the design of Starburst, and made the LWA expansion economical, alongside a dozen other things. Larry D'addario has gotten to the bottom of more problems than I can count. Michael Eastwood, Marin Anderson, Stephen Bourke and I kept each other sane while turning the LWA into a working instrument (including literally dragging fiber optic cable miles across the desert! I can also thank Dónal Ó Suilleabháin for teaching me the importance of computer security (lessons learned: do not trust your friends with your unlocked desktop), and Scott Barenfeld – the not-so-innocent

[1] *way* too much power.

bystander for many a prank, and always reliable for a firm slap to the back of the head!

Lastly, I want to thank my family and close friends for supporting me the entire way. Anna, helping me check the gravitational constant with a free-fall ballistic measurement. Jeremie, because it turns out that computer vision algorithms work as well for astronomical images as in Disney customers. Tamara, my most charming rock climbing counter-weight. Billy, laser-cutting a part or two late into the night for Starburst with me. My sister, Morgan, with her patience and understanding. My brother, Nathan, who inspired me to transfer back to Georgia Tech–a turning point in my life. And of course, my mother and father, two of the most honorable, diligent and supportive people I know, and the best role models I could hope for. I've faced many challenges thus far in life and fared extremely well–an overwhelmingly large reason for this is all of your advice, examples, and encouragement. I simply cannot express all of my gratitude in words.

ABSTRACT

In the past decade, there has been a revolution in radio-astronomy signal processing. High bandwidth receivers coupled with fast ADCs have enabled the collection of tremendous instantaneous bandwidth, but streaming computational resources are struggling to catch up and serve these new capabilities. As a consequence, there is a need for novel signal processing algorithms capable of maximizing these resources. This thesis responds to the demand by presenting FPGA implementations of a Polyphase Filter Bank which are an order of magnitude more efficient than previous algorithms while exhibiting similar noise performance. These algorithms are showcased together alongside a broadband RF front-end in Starburst: a 5 GHz instantaneous bandwidth two-element interferometer, the first broadband digital sideband separating astronomical interferometer. Starburst technology has been applied to three instruments to date.

Wielding tremendous computational power and precisely calibrated hardware, low frequency radio telescope arrays have potential greatly exceeding their current applications. This thesis presents new modes for low frequency radio-telescopes, dramatically extending their original capabilities. A microsecond-scale time/frequency mode empowered the Owens Valley Long Wavelength Array to inspect not just the radio sky by enabling the testing of novel imaging techniques and detecting overhead beacon satellites, but also the terrestrial neighborhood, allowing for the characterization and mitigation of nearby sources of radio frequency interference (RFI). This characterization led to insights prompting a nanosecond-scale observing mode to be developed, opening new avenues in high energy astrophysics, specifically related to the radio frequency detection of ultra-high energy cosmic rays and neutrinos.

Measurement of the flux spectrum, composition, and origin of the highest energy cosmic ray events is a lofty goal in high energy astrophysics. One of the most powerful new windows has been the detection of associated extensive air showers at radio frequencies. However, all current ground-based systems must trigger off an expensive and insensitive external source such as particle detectors - making detection of the rare, high energy events uneconomical. Attempts to make a direct detection in radio-only data have been unsuccessful despite numerous efforts. The problem is even more severe in the case of radio detection of ultra-high energy neutrino events, which cannot rely on in-situ particle detectors as a triggering mechanism. This thesis combines the aforementioned nanosecond-scale observing mode with real-time,

on-FPGA RFI mitigation and sophisticated offline post-processing. The resulting system has produced the first successful ground based detection of cosmic rays using only radio instruments. Design and measurements of cosmic ray detections are discussed, as well as recommendations for future cosmic ray experiments. The presented future designs allow for another order of magnitude improvement in both sensitivity and output data-rate, paving the way for the economical ground-based detection of the highest energy neutrinos.

PUBLISHED CONTENT AND CONTRIBUTIONS

Monroe, R. et al. (Submitted 2018[a]). "RFI Mitigation at Nanosecond Timescales". In: *Publications of the Astronomical Society of the Pacific*.
The author of this thesis designed the high time resolution capture system, performed all analysis and wrote the manuscript.

– (Submitted 2018[b]). "Self-triggered radio detection and identification of cosmic air showers with the OVRO-LWA". In: *Nuclear Instruments and Methods*.
The author of this thesis designed all firmware and software, performed all analysis and wrote the manuscript.

– (Submitted 2018[c]). "Starburst: A Flexible, RFI-Mitigating, Low-Cost Digital Backend Design for Radio Astronomy Applications". In: *Monthly Notices of the Royal Astronomical Society*.
The author of this thesis designed and tested the RF and digital systems, wrote the manuscript and participated in the field demonstration. JRV participated substantially in the project, particularly with regards to the field demonstration, as well as several parts of the project not covered in this chapter.

Monroe, R. (2016). "Improving the Performance and Resource Utilization of the CASPER FFT and Polyphase Filterbank". In: *Journal of Astronomical Instrumentation* 5, 1641002-262, pp. 1641002–262. DOI: 10.1142/S2251171716410026.
The author of this thesis is the sole author and contributor of this work.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 A new window into the universe with radio astronomy

Radio-astronomy is a powerful way to observe the universe. Spanning several decades (from 10 MHz – limited by the opacity of the ionosphere, to 1 THz – limited by atmospheric absorption and the fundamental limits of coherent amplifiers), it provides a window for a legion of science cases otherwise inaccessible. Popular instrument classes include spectrometers – single antenna instruments which measure the spectrum of a source (often equipped with broadband spectrometers), and arrays which use coherent analysis of the signals from groups of spatially distributed antennas to perform higher resolution. Currently, radio astronomical observations achieve the highest resolution astronomical observations of any performed by humans by several orders of magnitude, which is achieved by using Very Long Baseline Interferometry to combine signals from antennas across the planet (Alef, 2004). The three most important figures of merit for radio telescopes are bandwidth, system temperature and spatial resolution. The noise from an observation scales as $1/\sqrt{B\tau}$, with B being bandwidth and τ being integration time, so the extremely weak nature of radio astronomical signals often requires large collecting areas and cooled receivers (less true at low ~100 MHz) frequencies, where the bright galactic emissions make the design of sky-noise limited room temperature amplifiers practical). Another avenue for producing more science is through increased bandwidth. Although it depends on the application, this typically provides either more signal to integrate across, more parameter space to be sensitive to, or access to additional spectral lines to observe. This was typically limited by sampling rate or input bandwidth but at lower frequencies, fractional receiver bandwidth was an additional concern.

With the field of radio astronomy enjoying the benefits of Moore's Law, the capabilities of modern instruments are still rapidly expanding. This kind of rapid expansion in capabilities is not a new thing: radio-astronomy is limited by the most expensive component in the signal chain of – feed, radio frequency (RF) processing, sampling, and digital processing. This results in something of an "arms race" as technology is pushed to meet the capabilities of other hardware in the signal chain. Recently, the maximum ADC sampling rate was a limiting factor – in order

to sample beyond the maximum practical sampling rate (at that time, ~3 GSPS), proposed instruments had to split the signal chain into multiple bands and process each independently. However, doing so typically increases system complexity, cost, and power consumption. Recently, broadband ADCs are reaching the market, with sampling frequencies ranging from 5 GSPS up to 26 GSPS (Hittite, 2018). Between the recent development of 4x bandwidth feeds and high-performance samplers, it's now time for digital hardware and DSP to play catch-up.

The history of radio astronomy can be said to have begun with Karl Jansky's spontaneous detection of radio emission from the Milky Way[1] in 1932 (Condon and Ransom, 2016). After being reassigned to another task, the field laid untouched until Grote Reber designed and built a parabolic dish in his backyard in 1937 and producing the first radio maps of the sky in 1941. The success of Reber, combined with the influx of RADAR specialists after World War 2, sparked increasing interest in the field. Research branched out into interferometry with the first applications using radio reflections against the ocean to implement a crude additive interferometer (Sullivan, 1991), which increases spatial resolution substantially. Subsequently, aperture synthesis techniques (and especially earth rotation aperture synthesis)[2] (Fomalont, 1973) were made practical by the increasing availability of computational power. Of special interest to this thesis is the invention of the first digital back-end by Weinreb (1963) – a technique which is used in almost every radio astronomical instrument today. Over time, earth-borne and space-borne instruments for both earth-observing satellites and astronomy have transitioned from mostly analog to having substantial digital back-end reliance.

Interferometry: a cornerstone of modern radio astronomy

One limitation of radio-telescopes is that of resolution. The maximum angular resolution of a telescope (typically the full-width half power of the beam) is limited by $1.22\lambda/D$, with D and λ being the diameter of the aperture and the observation wavelength, respectively (R. Thompson, Moran, and Swenson, 2001). At low frequencies, λ can be as large as tens of meters (limited by the ionosphere). Since the largest steerable single dishes are about 100 m (Prestage et al., 2009), this is quite a problem. At radio frequencies, the standard solution is to use multiple antennas, sample their individual signals, and coherently combine them in order to emulate a

[1] This happened at Bell Labs, apparently the same as every other electronics discovery around that time.

[2] Which later won Ryle and Hewish the Nobel prize.

single larger dish in some way – this is called aperture synthesis.

Monochromaticity is an inherent assumption in interferometry (Zernike, 1938). In order to prevent bandwidth smearing (i.e., chromatic effects), processing sufficiently narrow bandwidths is mandatory; for reasons both scientific (investigating spectra) and practical, it is desirable to measure these correlations as a function of frequency. Initial instruments did this by measuring the cross-correlation of each pair of signals as a function of time and subsequently using a Fourier Transform (FT) to convert to the frequency domain – today, known as an XF correlator (X for multiply, F for Frequency). The method typically used today, however, takes advantage of the associativity of the FT and cross-correlation to perform the FTs first – reducing the number of required FTs from $\sim N^2$ to just N. Having the frequency transform operate directly on a voltage time-series allows for the replacement of the FT with a Polyphase Filterbank (discussed later), which has better RFI mitigation properties. On the other hand, not all multiplies are created equal – although the FX correlator (operating on the complex-valued output of a FT) has fewer mathematical operations, the XF correlator typically operates directly on (narrower and real-valued) ADC samples and therefore has simpler operations. Multiplying two 2-4 bit integers, such as those provided by an ADC, requires far less silicon than doing the same for the 4-8 bit complex-valued multiplications that must be performed after a frequency transform. For this reason, some specialized FPGA and ASIC correlators choose the XF architecture, especially if the systems are in favorable RFI environments.

1.2 Computational hardware for radio-astronomy back-ends

Most radio-astronomy back-ends look rather similar. An RF signal is provided by the telescope, receives initial analog signal processing (down-conversion, filtering, or amplification, for instance), and is digitized and processed before being sent to a PC for storage. Depending on the scale of the project, that digitization and processing can take many forms. There is a trade-off here between ease of programming (or equivalently, up-front cost), and performance per unit dollar or power. Here, these will arguably be presented in order of increasing performance:

Central Processing Units (CPUs)

The CPU is the workhorse of the consumer and commercial computing world. A CPU is presented with a sequence of instructions, each of which performs an elementary operation (such as "add", "multiply", and/or "store", although more

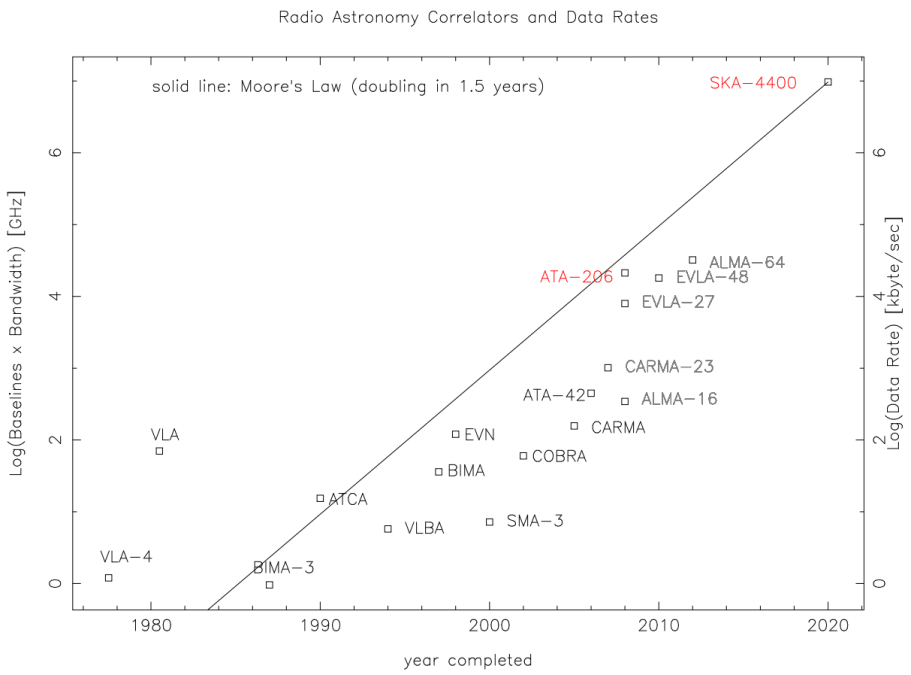


Figure 1.1: Bandwidth of radio astronomy correlators. Instruments in red have not yet been constructed. Figure credit: Mel Wright (2018)

sophisticated instructions exist). This set of instructions is shockingly versatile, and therefore any program can be written on a CPU. Most CPUs have ~MB sized on-chip cache and sophisticated prediction algorithms in order to maximize the efficiency and clock rate of their 1~4 cores. They are optimized for running a few threads extremely quickly, a design choice which is made to permit low-effort general-purpose software development. Because CPUs are typically used in the context of an operating system which time-multiplexes many programs, CPUs can perform task switching on ~ns timescales. CPUs are used for software VLBI correlators, which strike the right combination of small data-rates but extraordinary complexity. In general however, their flexibility and ease of use comes at a performance price: as we will see, adding structure to a computing task allows it to be performed far more efficiently than the naive CPU implementation.

Graphics Processing Units (GPUs)

GPUs impose structure on a problem in order to extract a performance gain. The devices remove the large caches and sophisticated prediction logic used in CPU, trading that functionality for a very large number of threads. While a typical CPU has ~4 threads, it is not uncommon to see a GPU with more than 2000 cores. In

order to experience these gains, GPU threads execute in groups (called "warps") of 32 threads – often known as Single Instruction Multiple Data. Naturally, this restricts the classes of problems that GPUs are good at. Additionally, the thread warp restrictions, as well as the weaker cache and prediction result in less efficient cores. Nonetheless, GPUs often exhibit ~100x performance gains on well-matched problems. Despite the thread warp restrictions, the space of problems which work well for GPUs is surprisingly large – including filtering, FFTs, cross-correlations, convolutional neural nets, sorting, counting, morphological operations, and more. Graphics Processing Units have a fascinating history. Originally, CPUs did not have the performance necessary to quickly draw all the graphics needed to human-viewable displays. This was especially true for the video gaming industry, which was constantly demanding higher resolutions and improved graphics. The high level of parallelism in these tasks (computing the color of each pixel was a largely independent task) drove the invention of the GPU. Eventually, as performance improved, some scientific tasks were mapped onto GPUs, starting with matrix multiplication and LU factorization in 2001 and 2005, respectively. NVIDIA released the Compute Unified Device Architecture (CUDA) library in 2006, which eventually became an industry dominator for GPU acceleration. There has been a recent explosion in GPU utilization through the advent of deep learning, which provides best-in-class machine learning for almost every task, and is trivially mappable to GPUs. As a consequence, the performance of GPUs has improved tremendously in recent years. The barrier to entry for GPU-based software is larger than CPU software. The debugging environment is inferior and it is harder to get data off of the devices (this has lessened in recent years, but is still a concern). GPUs excel when there are at least 32 independent data streams to process, which matches the problem well to the warp architecture. Most of the algorithms listed above only really experience great performance gains at large numbers of inputs or long filter/transform/data lengths. For this concern as well as the debugging issue, radio-astronomy projects tend to start using GPUs when the large-ish data-rates offset the pain of development.

Field Programmable Gate Arrays (FPGAs)

At the end of the day, you don't have a project if you can't get your data out of an RF cable. Any engineer who works with radioastronomy applications is painfully aware of the fact that commercial-off-the-shelf (COTS) hardware is unable to read the custom interfaces provided by high-speed RF ADCs and DACs. Field Programmable Gate Arrays (FPGAs) are almost the only class of product which can

perform this task. In fact, this task of interfacing otherwise incompatible hardware (often known as "glue logic" that ties a circuit together) was the original commercial value of FPGAs! FPGAs are a network of lookup tables (LUTs) which can implement general-purpose logic, as well as specialized hardware for mathematics, high-density memory, and off-chip communication. All of this hardware is connected with a switching network. By configuring these hardware elements and switching network, FPGAs can be re-programmed on ~second timescales. Because FPGAs have tremendous parallelism but very little on-chip memory (~10 Mb on the largest devices), they are optimal for parallelizable streaming data applications and applications which are sensitive to power efficiency. FPGA development is painfully slow and error-prone, but for well-matched applications, FPGAs experience a factor of 40~200 gain in power efficiency as compared to CPUs. FPGAs also have a restricted space of problems for which they are efficient, but it is different from that of GPUs. Specifically, since FPGAs perform all tasks in parallel, "if" statements are often computed by simply computing both options of the "if" statement and then choosing the appropriate result. Additionally, the on-chip memory restrictions of FPGAs are combined with difficult access to off-chip memory (such as DDR or QDR ram). For this reason, FPGAs are best suited to classical signal processing problems such as filter-banks, cross-correlation, or other signal processing applications. FPGAs have considerable commercial value: they exist in most of the large Internet backbone hardware, as well as used for many hardware prototyping applications. Development for FPGAs is relatively punishing. Unlike CPU and GPU code, an FPGA developer is effectively designing hardware. Although this hardware can be simulated to some extent in advance, extracting new information from a design at run-time requires recompiling the entire device, a process which took over five hours for the Starburst project (covered in Chapter 4). On top of this, FPGA designs typically must run at a certain clock rate to perform – in radio-astronomy contexts, this typically drives ADC sample rate and therefore maximum bandwidth. Achieving a sufficient clock rate involves a complicated interplay between logic design, physical placement of components on the FPGA chip, and routing interconnect between these components. A good tutorial for improving performance in a CASPER context (which generalizes well to most FPGA work) can be found at (Monroe, 2012). In short, FPGA designs are unpleasant to work with, and the source of many scheduling revisions. Nonetheless, they are present in almost every radio-astronomy task because they are practically mandatory for interfacing with high-performance ADCs. Two large correlators which feature FPGAs are WIDAR at the VLA and

CHIME.

Application Specific Integrated Circuits (ASICs)

Not the shoes! Although FPGAs enjoy performance gains over CPUs, there are still performance benefits to be extracted beyond their paradigm. Specifically,

- FPGAs ship with a standardized load-out of on-chip hardware, which is both imperfectly matched to the target application, and not in the optimal locations on-chip for the target application
- FPGAs are re-programmable – this adds otherwise unnecessary routing resources and switching matrices which could simply be hard-wired in a chip which was matched to the task
- FPGAs are re-programmable – in high radiation conditions, such as Low Earth Orbit, Single Event Upset events can change this configuration, altering the behavior of the device.

The answer to these three problems are Application Specific Integrated Circuits (ASICs): fully custom silicon that can perform a mix of RF and digital tasks on a single chip. However, ASICs are tremendously expensive for design and initial production, so their use is restricted to projects operating at the largest of scales, or space-borne projects with extremely tight power and/or radiation tolerance requirements. The Atacama Large Millimeter Array (ALMA) is one of the only recent correlators which uses ASICs. As alluded to above, the "best" configuration is experiment-dependent. Smaller experiments are more sensitive to man-hours (favoring GPU and CPU-centric implementations), whereas larger experiments care more about hardware cost and power consumption, which favors FPGAs and (in extreme cases) ASICs. Even in small-scale projects, however, the use of some kind of FPGA or ASIC (typically FPGA) is almost mandatory, at the very least for interfacing with high-performance ADCs and DACs (where they have a monopoly).

1.3 The community-driven tooling effort

Despite their ubiquity, using FPGAs typically implies working with custom hardware and firmware – an expensive prospect for low-volume projects such as most scientific research projects. The Collaboration for Astronomy Signal Processing and Electronics Research (CASPER) project largely mitigated this problem (Parsons et al., 2008;

Hickish et al., 2016). CASPER developed open-source Xilinx FPGA-based hardware platforms (current production hardware includes ROACH-2, SKARAB and SNAP2), which have become a community standard. In addition, they developed libraries which perform many common tasks needed by radio-astronomers and other scientists, as well as tool-flows which automate common tasks such as interfacing with external hardware. The libraries are extremely extensive and include communication interfaces (such as to ADCs, DACs, networking hardware and the host operating system), filters, windowing operations, frequency transforms, reordering elements, and accumulators. With ~45 active projects using their hardware and firmware, CASPER is a cornerstone of the radio astronomy community.

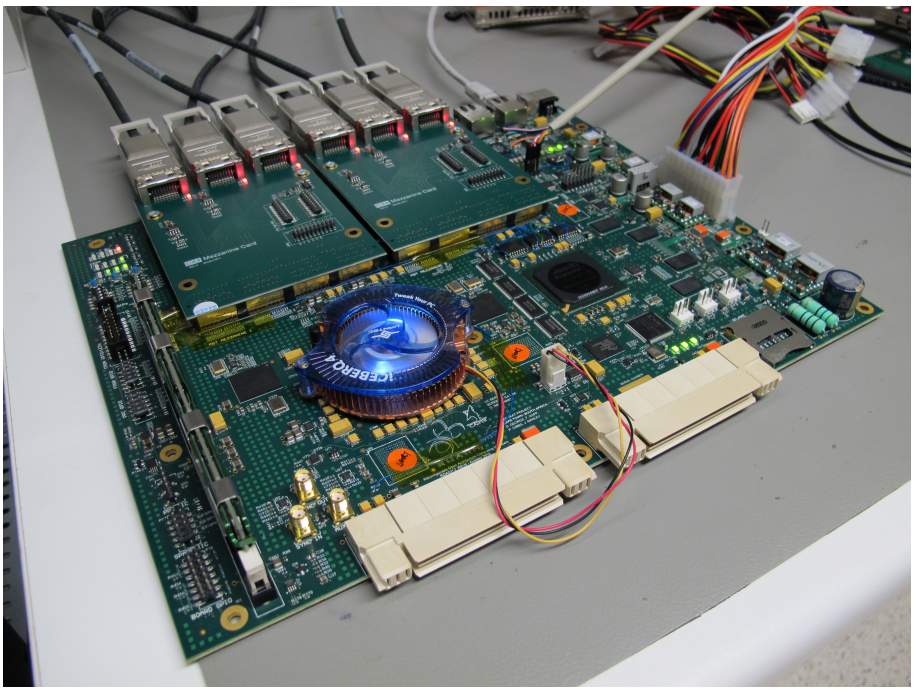


Figure 1.2: The ROACH2 is a CASPER-supported open source FPGA platform for radio astronomy applications.

CASPER uses Xilinx System Generator (implemented in MATLAB's Simulink) as their user-space firmware development environment. Unlike Verilog or VHDL, the environment is purely schematic based – a considerable departure from the industry standard hardware description languages of Verilog and VHDL. System Generator allows the user to compile most designs into a netlist which can be inserted into an arbitrary Xilinx-based design. With tight integration into MATLAB, System Generator is very effective at designing and debugging math-intensive signal processing tasks, features which are traded off with the platform's low stability and

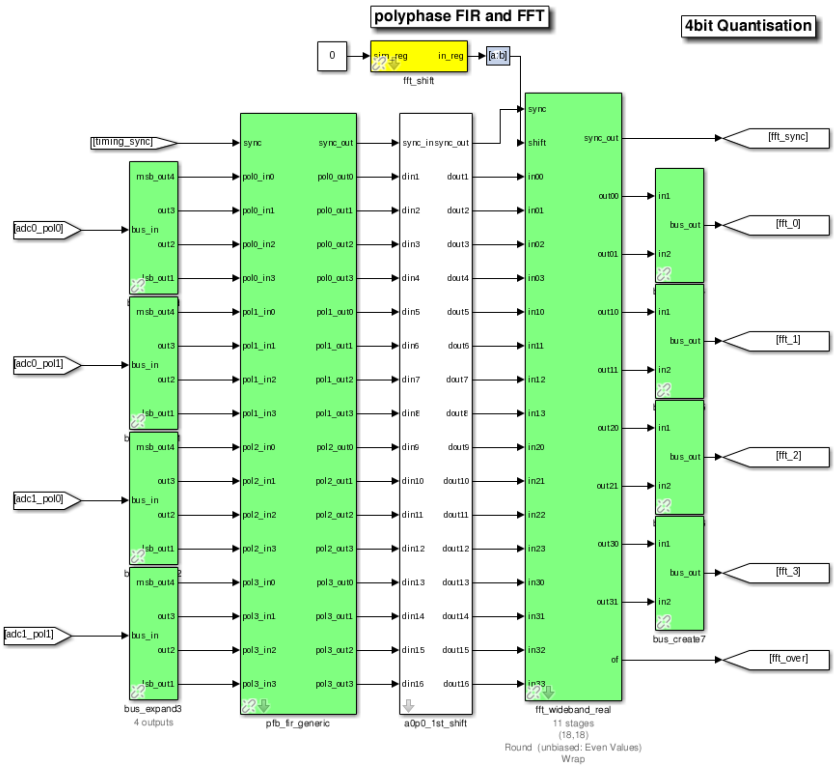


Figure 1.3: The schematic-style design environment used in Xilinx System Generator. Low-level primitives (functionally equivalent to structural HDL) can also be drawn procedurally in MATLAB.

limited feature set. Nonetheless, every work done for this thesis includes a System Generator design somewhere.

Frequency Transforms

Most interesting astronomical signals present as colored Gaussian noise. Depending on the temporal and frequency resolution needed, it is a common strategy to break the RF signal into a number of equally-spaced frequency bins, and integrate the power received in each bin over some desired temporal window. The integration time is generally driven by the properties of the signal chosen (such as the desired spatial resolution of Earth-observing mapping experiments, rates of dispersion for pulsar measurements, or variability of RFI sources).

Naively, this is can be thought of as a filter-bank, consisting of N Mixers followed by N identical low-pass filters and down-samplers, the collection of which is called a Discrete Fourier Transform (A. Oppenheim, 2009). However, for most practical applications, this $O\left(N^2\right)$ operation is implemented as a divide-and-conquer

algorithm that runs in $O(N \log N)$ time. These are called Fast Fourier Transforms (FFTs)– invented by Cooley and Tukey (1965), and a cornerstone of communications and signal processing. Henceforth, "FFT" will be used interchangeably for any discrete-time Fourier Transform.

As will be seen in Section 1.6, the frequency response of FFTs is inadequate for many sources of RFI. There are a few ways that this can be handled. For a given FFT length, there is a trade-off between channel width and out-of-band rejection. This can be performed by *windowing*, which minimizes discontinuities at the edges of the dataset. Windowing, however, broadens the main-lobe of the channel and produces a degradation in poor noise bandwidth. Intuitively, this makes sense: many samples are being down-weighted and this un-even treatment of samples is equivalent to a sub-optimal use of information in the presence of white Gaussian noise.

Intuitively, the issue with the FFT is insufficient window length. A Weighted Overlap Add (WOLA) FFT mitigates this problem by extending the analysis window by some integer factor T. In order to retain the same frequency resolution, this extended window is diced into segments of equal length to the original analysis window. Each of these is FFT'ed and finally summed. Roughly speaking, this gives you excellent ~50dB adjacent channel power rejection and a flat channel pass-band, without sacrificing much in terms of noise bandwidth. However, WOLA FFTs are computationally roughly T times more expensive. By taking advantage of the commutability of the FFT, the addition can be moved before the FFT. Called a Polyphase Filter Bank (Price, 2016; Vaidyanathan, 1992; Chennamangalam, 2011), this is computationally efficient for reasonable window scaling factors ($T < 8$ or so). The PFB still suffers from an inferior temporal response to an FFT, but this is not typically a major concern for most astronomy applications.

1.4 Currently deployed large radio experiments

Having discussed many of the important factors in correlator design leads us to cover currently active systems.

- "Low-Frequency Arrays": The currently constructed systems all use FX correlators, with an FPGA-centric F-engine and (with the exception of LOFAR) a GPU-centric X-engine.
  - The Owens Valley Long Wavelength Array (OVRO-LWA) makes instantaneous all-sky maps at 30 ~ 88 MHz. OVRO-LWA was the first

heterogeneous FPGA-F/GPU-X correlators (Kocz et al., 2015). The decision was also driven by the slow development times for FPGA correlators, as well as team expertise in designing peak performance GPU algorithms, enabling gains of a factor of a couple in computing efficiency for the GPU side.

- The Canadian Hydrogen Intensity Mapping Experiment (CHIME) (Denman et al., 2015) is a large drift-scan instrument with the goal of hydrogen mapping. Needing to iterate quickly on algorithms, they chose to implement their X-engine in GPUs.
- The Murchison Widefield Array (MWA) is an array based in Australia. Capable of large surveys of the Southern sky, MWA has many science objectives (Tingay et al., 2013), including the Epoch of Reionization (EoR), transient searches and space weather. Having relatively long baselines and antenna counts compared to previous instruments, the MWA performs analog beam-forming on tiles of 16 antennas, subsequently doing the standard correlation procedures with a GPU X-engine (Ord et al., 2015).
- The Hydrogen Epoch of Reionization Array (HERA) is a 350 element, 200 MHz bandwidth interferometer (DeBoer et al., 2017) designed purely to measure the EoR signal. With the hopes that it will ease calibration, HERA places their antennas into a grid pattern, resulting in as many repeated baselines as possible, a technique known as Redundant Calibration (Liu et al., 2010). They also use a FPGA (CASPER SNAP, specifically)-based F-engine and a GPU X-engine.
- The Low Frequency Array (LOFAR) (van Haarlem et al., 2013) spans much of Europe with 47 stations. The F-engines are co-located with the antennas for each of the 47 stations. Each produces one or more beams, which are then transmitted to a central processing area via Ethernet (Gunst and W Kant, 2018). At that location a CPU performs a multitude of tasks, including implementation of their X-engine for correlation and imaging (Romein et al., 2010).
- Upon completion, the Low Frequency Aperture Array of the Square Kilometer Array (SKA) is going to be the largest low-frequency array in existence (*The Square Kilometer Array* 2011). A number of the previous instruments - LWA, MWA, HERA - were technology demonstrators for

this all-singing, all-dancing instrument. Similar to LOFAR in a sense, SKA will produce local beam(s) from individual stations, and then cross-correlate them at a central station.

- The Very Large Array (VLA) was completed in 1980, and was the most powerful radio-telescope array of its time (A. R. Thompson et al., 1980; Napier, A. R. Thompson, and Ekers, 1983). After nearly three decades, the system was given a considerable upgrade. Although there were revisions to receivers and signal transport (using optical fiber instead of waveguides), one of the largest upgrades was an increase in instantaneous bandwidth by a factor of 80 (Perley et al., 2009) – improving sensitivity by a factor of almost 9 purely through the back-end! A departure from other systems discussed here, EVLA uses a XF correlator. I suspect that this is because EVLA required backwards compatibility with the VLA, which was created before FX correlators had been invented, in 1984 (for scale applications) (R. Thompson, Moran, and Jr., 2001; Chikada, Ishiguro, Hirabayashi, Morimoto, Morita, Miyazawa, et al., 1984; Chikada, Ishiguro, Hirabayashi, Morimoto, Morita, Kanzawa, et al., 1987)

1.5 Cosmic Rays and the Nanosecond Universe

Cosmic rays are extremely high energy particles mainly originating from outside our solar system. Typically protons or iron nuclei, these relativistic particles strike Earth's atmosphere at ~4km altitude, causing an "air shower" of progressively lower energy particles – mostly photons and charged particles, some of which strike the Earth's surface.

The radio emission comes from two sources. The air shower includes electron-positron pairs, traveling in approximately the direction of the incident cosmic ray: $\overrightarrow{v}$ at nearly the speed of light. These are acted on by the $q\overrightarrow{v} \times \overrightarrow{B}$ Lorentz force, due to Earth's geomagnetic field $\overrightarrow{B}$. The oppositely-charged electrons and positrons are accelerated in opposite directions, resulting in a net current which is very nearly parallel to $\overrightarrow{v} \times \overrightarrow{B}$. This produces synchrotron radiation with a magnitude which scales with the angle between $\overrightarrow{v}$ and $\overrightarrow{B}$, and location-independent polarization parallel to the Lorentz force. Producing about 99% of the emitted power (Huege and Besson, 2017), this is typically called the "Geomagnetic Effect". Scholten, Werner, and Rusydi (2008) showed that the maximum speed of these electrons (and therefore, the strength of the geomagnetic effect) is limited by collisions with atmospheric

constituents, and therefore roughly inversely proportional to atmospheric density.

The remaining 1% of the power is produced by the so-called "Askaryan Effect". Between Compton, Moeller and Bhabha scattering, a $10 \sim 20\%$ surplus of electrons is established in the shower front. These electrons repel each other, producing a radial current directed inward toward the shower axis, resulting in a location-dependent polarization. Since the component of velocity in parallel with the shower axis is very large and unaffected by atmospheric density (Scholten, Werner, and Rusydi, 2008), the Askaryan Effect is likewise density-independent. In both cases, because the emission is coherent at wavelengths larger than the size of the emitting system, the energy radiated in RF scales as roughly the number of particles squared (or equivalently, e^2) for sufficiently low frequencies $\nu < O\,(100\,MHz)$ (Scholten, Werner, and Rusydi, 2008). Nonetheless, the signals can be picked up from roughly $2 \sim 500$ MHz.

Both emission sources are highly beamed at $\sim 1^\circ \sim 2^\circ$. For events coming at high elevation angles, this implies a size scale on the ground of ~100m, but that is larger for events at lower elevation angles. The electric fields from the two emission mechanisms sum, creating an asymmetric resultant electric field. Because the geomagnetic effect is stronger at lower density while the Askaryan Effect remains unchanged, the power distribution of the emitted RF signature is more symmetric at lower atmospheric densities: this is confirmed by observations (Schellart, Buitink, et al., 2014). Since heavier cosmic ray primaries have a larger cross-section and strike at a lower atmospheric density (Müller et al., 2018), this symmetry (or lack thereof) can be used to measure X_{max} (the density of shower maximum), and therefore the atom that produced the cosmic ray event. Huege (2017) uses a more direct approach, directly estimating the cosmic ray primary from the power distribution on the ground – a technique which is more effective at lower elevation angles, where measuring X_{max} is more ill-conditioned.

An interested reader can read the excellent review and thesis of (Huege, 2016; Schröder, 2012), and references therein.

Although early detection of the radio signature of cosmic rays was demonstrated by Jelley et al. (1965), radio detections expanded briefly, but hit a lull and were not really picked up again until the 2000s. This was in large part due to the challenges with calibrating radio instruments back then, combined with the success of Fluorescence Detectors (Bergeson, Boone, and Cassiday, 1975). By approximately 2000, a combination of better understanding of the EM emission mechanism and improved

digital hardware performance combined with a need to read particle mass to make electronics appear more compelling.

The first two modern cosmic ray experiments were set up roughly simultaneously. CODALEMA (Ravel et al., 2004) equipped an existing radio telescope with particle detectors and a readout system. Multiple follow-ons upgraded the system from 11 antennas to 23 (CODALEMA-2) and 57 (CODALEMA-3) (Belletoile, 2011), the latter of which self-triggers on the RF signature and is in operation today. LOPES (which has the dubious distinction of operating in one of the worst possible RFI environments for RF cosmic ray detection, featuring welding, computers and a particle accelerator) equipped an existing particle detector array with radio-telescopes (Falcke et al., 2005), the goal being to demonstrate that the antennas could be used to detect cosmic rays at LOFAR. The experience from LOPES has resulted in LOFAR's tremendous results (Corstanje et al., 2011; Schellart, Nelles, et al., 2013), achieved by dual-using their array as a cosmic ray observatory. LOFAR is currently the largest radio-astronomy array to be simultaneously used as a cosmic ray instrument. TREND (Ardouin et al., 2011) performs RF self-triggers in China. Additionally, Tunka Radio Extension (Bezyazeekov et al., 2015) was seeking to cross-calibrate fluorescence and radio measurements. The Pierre Auger Observatory makes a very substantial effort to detect UHECRs, housing fluorescence and scintillation detectors and covering about $17km^2$. Co-located is the Auger Engineering Radio Array (AREA) (The Pierre Auger Collaboration et al., 2015) which demonstrates radio detection. An innovative balloon-borne instrument, ANITA (Hoover et al., 2010), looks for the back-scatter of UHECR events against the Antarctic ice – a clever way to deal with the low fluxes of UHECRs.

There are a few mechanisms by which cosmic rays are currently detected:

• Water Cherenkov Detectors or Surface Detectors (SD) are simply tanks of water placed on the ground. Those particles from cosmic ray events which reach the surface of the earth can strike the water contained within, producing Cherenkov radiation which can be detected. These detectors have a low false-positive rate, but are ineffective at detecting cosmic ray events which occur at low elevation angles: the particles have more atmosphere to travel through, and therefore don't make it to the detectors. Most radio detectors (including LOFAR) use these to detect cosmic ray events and trigger their RF voltage dumps.

• Fluorescence Detectors (FD): telescopes which measure the amount of Nitrogen fluorescence along the cosmic ray's path of travel. X_{max} can be read directly from the

fluorescence data, making these telescopes very attractive. However, because they operate by telescope observation, Fluorescence Detectors only work on dark nights and are subject to atmospheric and weather conditions (Pierre Auger Collaboration, 2015). The Pierre Auger Observatory (The Pierre Auger Collaboration, 2015) is only able to use its FDs about 15% of the time.

• Radio Frequency detectors, which use the impulsive radio signature to detect cosmic rays, and discriminate them from RFI using metrics such as the direction of arrival (in arrays). RF detection can measure everything achievable by FDs, but with full duty cycle. However, RF detectors must combat an (often very challenging) impulsive RFI environment.

While the lower energy cosmic rays are believed to come from supernova remnants (Blasi, 2013), the source of the highest energy cosmic rays remains an open question (current theories include active galactic nuclei and galaxy mergers). Unfortunately, measuring the direction of arrival for cosmic ray events is unhelpful: magnetic fields experienced on the path of travel redirect the charged particle, destroying the source information. The current theories of high energy Cosmic Ray origin and sources can be constrained by knowing the mass composition of these events – a task well-suited to radio observation. Additionally, the flux spectrum of cosmic rays scales with roughly E^{-3}, meaning that high energy events are very rare. This drives a need for good spatial coverage and high duty cycle observation: re-purposing existing radio-telescope arrays would provide plenty of collecting area and spatial coverage at a relatively small additional cost. Radio arrays have an added benefit, provided that reliable low-elevation-angle detection can be achieved: cosmic rays arriving at low elevation angles generally strike the atmosphere farther from the detector, giving the powerful signature more distance with which to spread out, and therefore requiring a lower spatial density of antennas. In comparison, FDs have a poor duty cycle, and SDs are both expensive and unable to observe at these low elevation angles.

One related scientific opportunity comes in the form of neutrinos. Because they interact with the rest of the universe so rarely, neutrinos provide a window into the otherwise opaque high-energy universe. The highest energy cosmic rays are largely suppressed by the Greisen–Zatsepin–Kuzmin limit[3], and the universe is opaque to the TeV and above energy photons (Halzen and Klein, 2010; Halzen and Hooper, 2002), making neutrinos one of the only options in this regime. Furthermore,

[3] Actually only mostly – why this is the case is an open question

neutrinos are believed to both be produced by an interaction between high energy cosmic rays with the Cosmic Microwave Background, as well as believed (in some models) to be directly produced by sources of high energy cosmic rays.

Naturally, the weak interaction between neutrinos and the rest of the universe makes them challenging to detect. Specifically, neutrinos produce no particle shower and therefore can only be detected by their radio signature. Thus far, no radio experiment has conclusively detected a neutrino event (Gorham et al., 2016; Abraham et al., 2009), making this a promising opportunity for a large radio array to produce new science. Several radio instruments (Martineau-Huynh et al., 2017) are being developed to tackle this problem, and the work presented in this thesis makes them look more feasible.

1.6 Radio Frequency Interference

The great challenge that radio detection of cosmic rays faces is Radio Frequency Interference. This is shared with the larger radio-astronomy community: the larger bandwidth enabled in part by fast compute comes with larger exposure to RFI. Astronomers are not the only users of the electromagnetic spectrum, and modern electronics emit (intentionally or otherwise) considerable interference. Scientists have protected bands of the electromagnetic spectrum in many jurisdictions: it is illegal for unauthorized users to transmit in those bands. Enforcement is difficult, however (producing antennas which can localize RFI sources at long wavelengths is especially challenging) and those bands typically are not wide enough to protect the entire frequency range desired. As a consequence, radio astronomers must coexist with terrestrial, airborne, and space-borne radio users. This problem has become more severe as time passes for three major reasons:

- The use of radio technology has increased (for instance, cellular communications, WI-FI, anti-vehicle collision RADAR)
- The bandwidth needs of radio astronomers have increased – electronics have gotten faster and cheaper, relative to the price of antennas.
- Broader band ADCs typically have fewer bits and therefore less dynamic range – which makes them more susceptible to RFI.

Most RFI flagging techniques take advantage of the fact that most astronomical sources are spectrally smooth and, time-invariant. Flagged data is ignored for

subsequent processing. For arrays, the direction of arrival can be used for RFI filtering as well, since RFI sources often come from a fixed or predictable direction.

Many common RFI sources cover a narrow frequency bandwidth ("narrow-band"). When possible, it is desirable to flag these sources after the frequency transform and integration. In order to get an adequate out-of-frequency-channel rejection, the aforementioned PFB is used for channelization. Afterward, non-smoothness in the time or frequency domains is often used as an indicator of RFI (in the absence of an astrophysical explanation of that behavior).

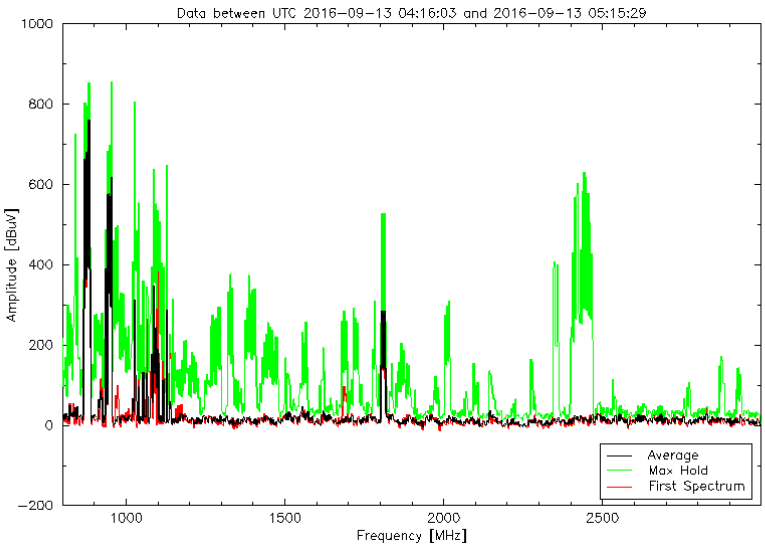


Figure 1.4: Radio Frequency Interference, as seen at the Australia Telescope Compact Array (ACTA, 2018). The power at ~1 GHz is likely cellular transmissions, whereas most of the power represented by the green trace (max-hold) is likely due to either low duty-cycle broadband impulses or saturation of some component of their signal chain.

For low duty-cycle RFI (as compared to the integration length), the kurtosis statistic (a measure of the tailedness of a distribution) has been used effectively for RFI detection. Although the kurtosis statistic is traditionally performed on real-valued signals, Spectral Kurtosis (Nita and Gary, 2010) is a variant with similar properties which works on complex-valued variables such as the output of a frequency transform. In general, kurtosis-class algorithms seem to work well when the RFI duty cycle is less than $\sim 1/N$, where N is the number of samples to be averaged across.

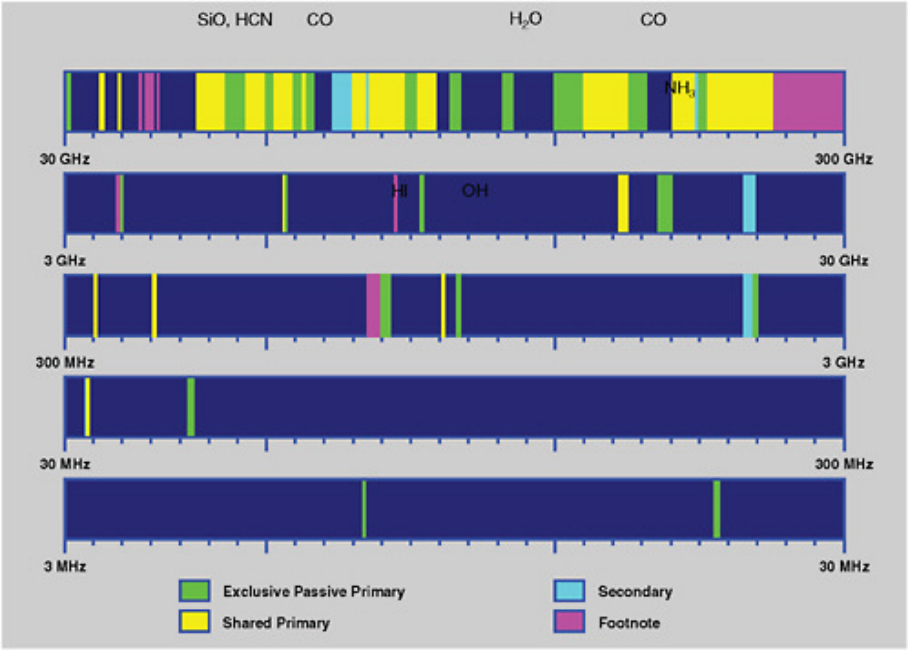


Figure 1.5: Frequency bands protected for radio astronomy in the United States (National Research Council, 2010). Some of these bands are at risk of re-assignment for commercial uses.

References

A. Oppenheim, R. Schafer (2009). *Discrete-time signal processing*. 3rd ed. Prentice-Hall, pp. 623, 723, 746.

Abraham, J. et al. (2009). "Limit on the diffuse flux of ultrahigh energy tau neutrinos with the surface detector of the Pierre Auger Observatory". In: *Physical Review D* 79.10, 102001, p. 102001. DOI: 10.1103/PhysRevD.79.102001. arXiv: 0903.3385 [astro-ph.HE].

ACTA (2018). *ATCA Users Guide*. URL: http://www.narrabri.atnf.csiro.au/observing/users_guide/html/atug.html.

Alef, W. (2004). "A Review of VLBI Instrumentation". In: *European VLBI Network on New Developments in VLBI Science and Technology*. Ed. by R. Bachiller et al., pp. 237–244. eprint: astro-ph/0412294.

Ardouin, D. et al. (2011). "First detection of extensive air showers by the TREND self-triggering radio experiment". In: *Astroparticle Physics* 34, pp. 717–731. DOI: 10.1016/j.astropartphys.2011.01.002. arXiv: 1007.4359 [astro-ph.IM].

Belletoile, A. (2011). "First results of the new Autonomous Antenna Array of the CODALEMA radio detection experiment". In: *International Cosmic Ray*

Conference 1, p. 231. DOI: 10.7529/ICRC2011/V01/0299. arXiv: 1109.3579 [astro-ph.IM].

Bergeson, H. E., J. C. Boone, and G. L. Cassiday (1975). "The Fly's Eye. A Novel Technique for Sensing Extensive Air Showers". In: *International Cosmic Ray Conference* 8, p. 3059.

Bezyazeekov, P. A. et al. (2015). "Measurement of cosmic-ray air showers with the Tunka Radio Extension (Tunka-Rex)". In: *Nuclear Instruments and Methods in Physics Research A* 802, pp. 89–96. DOI: 10.1016/j.nima.2015.08.061. arXiv: 1509.08624 [astro-ph.IM].

Blasi, P. (2013). "The origin of galactic cosmic rays". In: *Astronomy and Astrophysics Reviews* 21, 70, p. 70. DOI: 10.1007/s00159-013-0070-7. arXiv: 1311.7346 [astro-ph.HE].

Chennamangalam, Jayanth (2011). *The Polyphase Filter Bank Technique*. https://casper.berkeley.edu/wiki/The_Polyphase_Filter_Bank_Technique.

Chikada, Y., M. Ishiguro, H. Hirabayashi, M. Morimoto, K. Morita, T. Kanzawa, et al. (1987). "A 320-MHz 1024-channel FFT cross-spectrum analyzer for radio astronomy". In: *Proceedings of the IEEE* 75.9, pp. 1203–1210. ISSN: 0018-9219. DOI: 10.1109/PROC.1987.13873.

Chikada, Y., M. Ishiguro, H. Hirabayashi, M. Morimoto, K. Morita, K. Miyazawa, et al. (1984). "A Digital FFT Spectro-Correlator for Radio Astronomy". In: *Indirect Imaging. Measurement and Processing for Indirect Imaging*. Ed. by J. A. Roberts, p. 387.

Condon, James and Scott Ransom (2016). *Essential Radio Astronomy*. Princeton University Press. URL: https://www.cv.nrao.edu/~sransom/web/xxx.html.

Cooley, James W. and John W. Tukey (1965). "An Algorithm for the Machine Calculation of Complex Fourier Series". In: *Mathematics of Computation* 19.90, pp. 297–301. ISSN: 00255718, 10886842. URL: http://www.jstor.org/stable/2003354.

Corstanje, A. et al. (2011). "LOFAR: Detecting Cosmic Rays with a Radio Telescope". In: *ArXiv e-prints*. arXiv: 1109.5805 [astro-ph.HE].

DeBoer, David R. et al. (2017). "Hydrogen Epoch of Reionization Array (HERA)". In: *Publications of the Astronomical Society of the Pacific* 129.974, p. 045001. URL: http://stacks.iop.org/1538-3873/129/i=974/a=045001.

Denman, N. et al. (2015). "A GPU-based Correlator X-engine Implemented on the CHIME Pathfinder". In: *ArXiv e-prints*. arXiv: 1503.06202 [astro-ph.IM].

Falcke, H. et al. (2005). "Detection and imaging of atmospheric radio flashes from cosmic ray air showers". In: *Nature* 435, pp. 313–316. DOI: 10.1038/nature03614. eprint: astro-ph/0505383.

Fomalont, E. B. (1973). "Earth-rotation aperture synthesis". In: *Proceedings of the IEEE* 61.9, pp. 1211–1218. ISSN: 0018-9219. DOI: 10.1109/PROC.1973.9247.

Gorham, P. W. et al. (2016). "Characteristics of Four Upward-Pointing Cosmic-Ray-like Events Observed with ANITA". In: *Physical Review Letters* 117.7, 071101, p. 071101. DOI: 10.1103/PhysRevLett.117.071101. arXiv: 1603.05218 [astro-ph.HE].

Gunst, Andre and Gideon W Kant (2018). "Signal Transport and Processing at the LOFAR Remote Stations". In: URL: https://tinyurl.com/y7gm6295.

Halzen, F. and D. Hooper (2002). "High-energy neutrino astronomy: the cosmic ray connection". In: *Reports on Progress in Physics* 65, pp. 1025–1078. DOI: 10.1088/0034-4885/65/7/201. eprint: astro-ph/0204527.

Halzen, F. and S. R. Klein (2010). "Invited Review Article: IceCube: An instrument for neutrino astronomy". In: *Review of Scientific Instruments* 81.8, pp. 081101–081101. DOI: 10.1063/1.3480478. arXiv: 1007.1247 [astro-ph.HE].

Hickish, J. et al. (2016). "A Decade of Developing Radio-Astronomy Instrumentation using CASPER Open-Source Technology". In: *Journal of Astronomical Instrumentation* 5, 1641001-12, pp. 1641001–12. DOI: 10.1142/S2251171716410014. arXiv: 1611.01826 [astro-ph.IM].

Hittite (2018). *3-BIT 26 GSPS ANALOG-TO-DIGITAL CONVERTER W/ OVERRANGE, INHIBIT, AND 1:2 DEMUX*. URL: http://www.analog.com/media/en/technical-documentation/data-sheets/hmcad5831.pdf.

Hoover, S. et al. (2010). "Observation of Ultrahigh-Energy Cosmic Rays with the ANITA Balloon-Borne Radio Interferometer". In: *Physical Review Letters* 105.15, 151101, p. 151101. DOI: 10.1103/PhysRevLett.105.151101. arXiv: 1005.0035 [astro-ph.HE].

Huege, T. (2016). "Radio detection of cosmic ray air showers in the digital era". In: *Physics Reports* 620, pp. 1–52. DOI: 10.1016/j.physrep.2016.02.001. arXiv: 1601.07426 [astro-ph.IM].

– (2017). "Radio detection of extensive air showers". In: *Nuclear Instruments and Methods in Physics Research A* 876, pp. 9–12. DOI: 10.1016/j.nima.2016.12.012. arXiv: 1701.02995 [astro-ph.IM].

Huege, T. and D. Besson (2017). "Radio-wave detection of ultra-high-energy neutrinos and cosmic rays". In: *Progress of Theoretical and Experimental Physics* 2017.12, 12A106, 12A106. DOI: 10.1093/ptep/ptx009. arXiv: 1701.02987 [astro-ph.IM].

Jelley, J. V. et al. (1965). "Radio Pulses from Extensive Cosmic-Ray Air Showers". In: *Nature* 205, pp. 327–328. DOI: 10.1038/205327a0.

Kocz, J. et al. (2015). "Digital Signal Processing Using Stream High Performance Computing: A 512-Input Broadband Correlator for Radio Astronomy". In: *Journal of Astronomical Instrumentation* 4, 1550003, p. 1550003. DOI: 10.1142/S2251171715500038. arXiv: 1411.3751 [astro-ph.IM].

Liu, Adrian et al. (2010). "Precision calibration of radio interferometers using redundant baselines". In: *Monthly Notices of the Royal Astronomical Society* 408.2, pp. 1029–1050. DOI: 10.1111/j.1365-2966.2010.17174.x. eprint: /oup/backfile/content_public/journal/mnras/408/2/10.1111_j.1365-2966.2010.17174.x/1/mnras0408-1029.pdf. URL: http://dx.doi.org/10.1111/j.1365-2966.2010.17174.x.

Martineau-Huynh, O. et al. (2017). "The Giant Radio Array for Neutrino Detection". In: *European Physical Journal Web of Conferences*. Vol. 135. European Physical Journal Web of Conferences, p. 02001. DOI: 10.1051/epjconf/201713502001. arXiv: 1702.01395 [astro-ph.IM].

Mel Wright (2018). *Radio Astronomy — Wikipedia, The Free Encyclopedia*. [Online; accessed 23-April-2018]. URL: http://casper.berkeley.edu/presentations/workshop_august08/slides/nsf_short.pdf.

Monroe, Ryan (2012). *Performance optimization for Virtex 6 CASPER designs*. https://www.dropbox.com/s/jnlnl0f7n78ogju/roach2_timing.zip?dl=1.

Müller, S. et al. (2018). "Impact of muon detection thresholds on the separability of primary cosmic rays". In: *Astroparticle Physics* 97, pp. 174–185. DOI: 10.1016/j.astropartphys.2017.11.005.

Napier, P. J., A. R. Thompson, and R. D. Ekers (1983). "The very large array: Design and performance of a modern synthesis radio telescope". In: *Proceedings of the IEEE* 71.11, pp. 1295–1320. ISSN: 0018-9219. DOI: 10.1109/PROC.1983.12765.

National Research Council (2010). *Spectrum Management for Science in the 21st Century*. Washington, DC: The National Academies Press. ISBN: 978-0-309-14686-9. DOI: 10.17226/12800. URL: https://www.nap.edu/catalog/12800/spectrum-management-for-science-in-the-21st-century.

Nita, G. M. and D. E. Gary (2010). "The generalized spectral kurtosis estimator". In: *Monthly Notices of the Royal Astronomical Society* 406, pp. L60–L64. DOI: 10.1111/j.1745-3933.2010.00882.x. arXiv: 1005.4371 [astro-ph.IM].

Ord, S. M. et al. (2015). "The Murchison Widefield Array Correlator". In: *Publications of the Astronomical Society of Australia* 32, e006, e006. DOI: 10.1017/pasa.2015.5. arXiv: 1501.05992 [astro-ph.IM].

Parsons, A. et al. (2008). "A Scalable Correlator Architecture Based on Modular FPGA Hardware, Reuseable Gateware, and Data Packetization". In: *Publications of the ASP* 120, p. 1207. DOI: 10.1086/593053. arXiv: 0809.2266.

Perley, R. et al. (2009). "The Expanded Very Large Array". In: *IEEE Proceedings* 97, pp. 1448–1462. DOI: 10.1109/JPROC.2009.2015470. arXiv: 0909.1585 [astro-ph.IM].

Pierre Auger Collaboration (2015). "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research A* 798, pp. 172–213. DOI: 10.1016/j.nima.2015.06.058.

Prestage, R. M. et al. (2009). "The Green Bank Telescope". In: *Proceedings of the IEEE* 97.8, pp. 1382–1390. ISSN: 0018-9219. DOI: 10.1109/JPROC.2009.2015467.

Price, D. C. (2016). "Spectrometers and Polyphase Filterbanks in Radio Astronomy". In: *ArXiv e-prints*. arXiv: 1607.03579 [astro-ph.IM].

Ravel, O. et al. (2004). "Radio detection of cosmic ray air showers by the CODALEMA experiment". In: *Nuclear Instruments and Methods in Physics Research A* 518, pp. 213–215. DOI: 10.1016/j.nima.2003.10.063. eprint: astro-ph/0306255.

Romein, John W. et al. (2010). "The LOFAR Correlator: Implementation and Performance Analysis". In: *SIGPLAN Not.* 45.5, pp. 169–178. ISSN: 0362-1340. DOI: 10.1145/1837853.1693477. URL: http://doi.acm.org/10.1145/1837853.1693477.

Schellart, P., S. Buitink, et al. (2014). "Polarized radio emission from extensive air showers measured with LOFAR". In: *JCAP* 10, 014, p. 014. DOI: 10.1088/1475-7516/2014/10/014. arXiv: 1406.1355 [astro-ph.HE].

Schellart, P., A. Nelles, et al. (2013). "Detecting cosmic rays with the LOFAR radio telescope". In: *A&A* 560, A98, A98. DOI: 10.1051/0004-6361/201322683. arXiv: 1311.1399 [astro-ph.IM].

Scholten, O., K. Werner, and F. Rusydi (2008). "A macroscopic description of coherent geo-magnetic radiation from cosmic-ray air showers". In: *Astroparticle Physics* 29, pp. 94–103. DOI: 10.1016/j.astropartphys.2007.11.012. arXiv: 0709.2872.

Schröder, F. G. (2012). *Instruments and Methods for the Radio Detection of High Energy Cosmic Rays*. DOI: 10.1007/978-3-642-33660-7. URL: https://www.astro.ru.nl/lopes/_media/publications/thesisfrankschroeder.pdf.

Sullivan III, W. T. (1991). "Some highlights of Interferometry in early Radio Astronomy". In: *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*. Ed. by T. J. Cornwell and R. A. Perley. Vol. 19. Astronomical Society of the Pacific Conference Series, p. 132.

The Pierre Auger Collaboration (2015). "The Pierre Auger Cosmic Ray Observatory". In: *ArXiv e-prints*. arXiv: 1502.01323 [astro-ph.IM].

The Pierre Auger Collaboration et al. (2015). "The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015)". In: *ArXiv e-prints*. arXiv: 1509.03732 [astro-ph.HE].

The Square Kilometer Array (2011). URL: https://www.skatelescope.org/wp-content/uploads/2011/03/SKA-Brochure.pdf.

Thompson, A. R. et al. (1980). "The Very Large Array". In: *Astrophysical Journal, Supplement* 44, pp. 151–167. DOI: 10.1086/190688.

Thompson, Richard, James Moran, and George Swenson Jr. (2001). *Interferometry and Synthesis in Radio Astronomy*. Wiley-VCH, pp. 290–291. ISBN: 0471254924. URL: https://www.amazon.com/Interferometry-Synthesis-Astronomy-Richard-Thompson/dp/0471254924?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0471254924.

Thompson, Richard, James Moran, and George Swenson (2001). *Interferometry and Synthesis in Radio Astronomy*. Wiley-VCH. ISBN: 0471254924. URL: https://www.amazon.com/Interferometry-Synthesis-Astronomy-Richard-Thompson/dp/0471254924?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0471254924.

Tingay, S. J. et al. (2013). "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies". In: *Publications of the Astron. Soc. of Australia* 30, e007, e007. DOI: 10.1017/pasa.2012.007. arXiv: 1206.6945 [astro-ph.IM].

Vaidyanathan, PP. (1992). *Multirate systems and filter banks*. Prentice Hall.

van Haarlem, M. P. et al. (2013). "LOFAR: The LOw-Frequency ARray". In: *Astronomy and Astrophysics* 556, A2, A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

Weinreb, Sander (1963). "A digital spectral analysis technique and its application to radio astronomy". PhD thesis. Massachusetts Institute of Technology. URL: https://dspace.mit.edu/handle/1721.1/11599.

Zernike, F (1938). "The concept of degree of coherence and its application to optical problems". In: 5, pp. 785–795.

Chapter 2

OPTIMIZED STREAMING FPGA ALGORITHMS FOR THE CASPER ARCHITECTURE

2.1 Introduction

As mentioned in the introduction, the open-source hardware and software developed by CASPER massively reduce the barrier-of-entry to FPGA signal processing for science applications. For this reason, CASPER hardware and software are heavily used in the radioastronomy community. However, these algorithms were developed with portability and ease-of-use in mind, rather than efficiency. While huge projects can bring the necessary resources to bear to produce efficient algorithms and well-matched hardware, the typical astronomer uses those provided by CASPER. This work focuses on bringing peak performance signal processing within the reach of smaller radio-astronomy experiments.

Algorithm design in FPGAs is somewhat different from that of CPUs. Existing computational resources in the FPGA are allocated to tasks which are relatively fixed for the duration of that program. Because there is a finite number of each type of FPGA resource (logic, memory, and arithmetic units), the total amount of processing achievable is limited by the resources consumed by each operation. The throughput of these tasks is then determined by the maximum clock rate of the slowest algorithm. Most radioastronomy applications process data in a full-duty-cycle, or "streaming" mode, meaning that an increased clock rate translates linearly to an increased maximum bandwidth. Because a faster and more resource-efficient algorithm results in more sophisticated processing options, there is tremendous value in improving FPGA-based DSP algorithms. Since many science DSP applications involve dividing a signal into spectral channels for further processing, most designs include a Polyphase Filter Bank (PFB), which consists of a Polyphase Finite Impulse Response filter, (e.g. Lyons, 2010) (hereafter denoted Polyphase-FIR, the name of this module in CASPER[1]), followed by a Fast Fourier Transform (FFT) (e.g. A. Oppenheim, 2009). These are very computationally intensive algorithms: in this author's opinion, representing a substantial portion of the FPGA utilization is a typical CASPER application. Furthermore, although the CASPER algorithms are a

[1] `https://casper.berkeley.edu/wiki/images/2/24/Casper\_memo\_pfb.pdf`

good match for the streaming architecture of the problem, they were designed with ease-of-use, flexibility, and generality as a priority, rather than performance. This makes these algorithms obvious targets for optimization. This paper will discuss optimizations resulting in a massive performance improvement compared to baseline CASPER designs. All designs referenced in this paper are available open-source on the CASPER community website, and are available from the author upon request.

2.2 Improved FFT Architectures

For wideband DSP tasks, it is often necessary to process samples at a rate greater than the maximum FPGA clock rate. In these situations, multiple ADC samples must be presented to the FPGA every clock cycle: this process is known as “demultiplexing”, or “demux” for short. Demultiplexing effectively trades a lower clock rate for higher resource utilization; the trade-off should be roughly linear. Since the highest practical clock rate on a Virtex 6 FPGA (speed grade-1) is ~325 MHz, and efficient implementations require a demux factor of 2^L, the optimal demux factor is typically

$$2^{ceil\left(\log_2\left(S/325\text{ MHz}\right)\right)}, \tag{2.1}$$

where S is the sample rate.

The Decimation in Time (DIT) FFT see A. Oppenheim, 2009, implements this divide-and-conquer algorithm by starting with smaller FFTs and combining them hierarchically into larger FFTs. In this way, the output of the s’th stage of a N-point FFT consists of $\frac{N}{2^s}$ FFTs, each of which are 2^s-points long. Because each of those FFTs accepts as an input every $\frac{N}{2^s}$’th input sample, CASPER-style FFTs implement a practical wideband FFT by separating the mathematical operations which operate on a single series of demultiplexed samples (the first $log_2 M$ FFT stages, which are calculating $M = {}^N/_L$-point FFTs), followed by $log_2(L)$ stages which do involve interaction. Internally, CASPER names these algorithms "fft_biplex" and "fft_direct", respectively. For the general case of a complex-valued input N-point FFT, with a demultiplexing factor of L, the "fft_biplex" modules will compute $L \cdot \frac{M}{2} \log_2 M$ “butterfly” operations, whereas "fft_direct" will compute $M \cdot \frac{L}{2} \log_2 L$ of these operations. Together, they have to perform the same number of butterfly operations as a single $N = LM$-point FFT:

$$M \cdot \frac{L}{2} \log_2 L + L \cdot \frac{M}{2} \log_2 M = \frac{N}{2} \log_2 N. \tag{2.2}$$

An additional concern is memory complexity: Delay Commutator FFTs such as those used in serial CASPER FFTs (He and Torkelson, 1998) must store $2^{\log_2(M)-S}$

complex-valued data points in each stage, for a total of N complex samples. Additionally, the data may be reordered at the output, undoing the bit-order-reversal produced by the FFT, requiring a further N complex samples. Therefore, the "naive" DIT requires space to store approximately $2N$ complex samples.

The typical input to a CASPER FFT is real-valued ADC samples. Since an N-point (complex-input) FFT can simultaneously perform two N-point (real-input) FFTs through the addition of some trivial trigonometric operations, CASPER uses this technique in the initial ("fft_biplex") FFTs, for a total $\frac{M}{2} \cdot \frac{L}{2} \log_2 L + L \cdot \frac{M}{2} \log_2 M$ butterflies. Because the initial L-point FFTs are performed at half size, they only require storage for N complex samples. However, the conjugate-symmetric negative frequency channels must be produced from the positive frequency channels, requiring a further $N/2$ samples, for a total memory requirement of $3N/2$ complex samples. From an algorithm implementation standpoint, the end of the M-point FFTs is a convenient breaking point because individual FFTs can be treated more independently, simplifying simulation and debugging. However, since this "dual real FFT in one complex FFT" trick is only applied to the first FFT module, the second FFT's efficiency is not improved. For FFTs with a large demux factor compared to their length, this second FFT is a dominant part of the computational complexity of the algorithm. This is the standard realization of wideband CASPER FFTs, and can be seen in Figure 2.1.

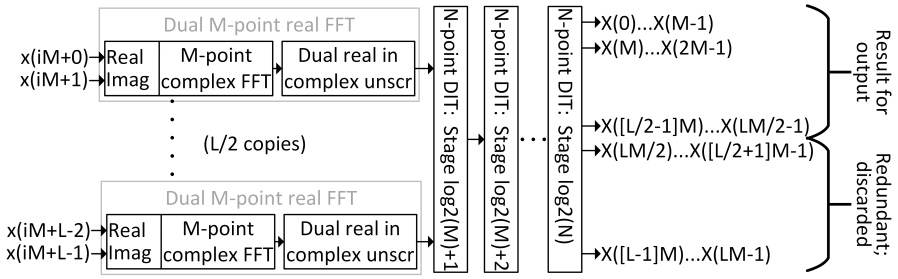


Figure 2.1: Traditional CASPER wideband FFT architecture.

Taking this idea to the logical conclusion, the "dual real FFT in one complex FFT" (W. Smith, 1995a) trick can be applied to the entire FFT. This results in two output FFTs, one representing the spectrum of the even input samples, and the other representing the "odd" input samples. A final DIT FFT stage synthesizes the non-redundant output spectrum (Figure 2.2). The first M-point FFTs retain the same complexity of $\frac{M}{2} \cdot \frac{L}{2} \log_2 L$, but the final stages ("fft_direct") is reduced to $\frac{L}{2} \cdot \frac{M}{2} \log_2 \frac{M}{2}$. The final DIT stage that was added to the end of the algorithm results

in the total computational complexity of

$$\frac{M}{2} \cdot \frac{L}{2} \log_2 L + L \cdot \frac{M}{4} \log_2 \frac{M}{2} + L \cdot \frac{M}{2}. \tag{2.3}$$

This represents a savings of $\frac{LM}{4}\left(\log_2 M - 1\right)$ butterflies, when compared to the original CASPER implementation. To produce a naturally-ordered FFT output with this algorithm requires the same memory operations, and therefore the same $3N/2$ memory space.

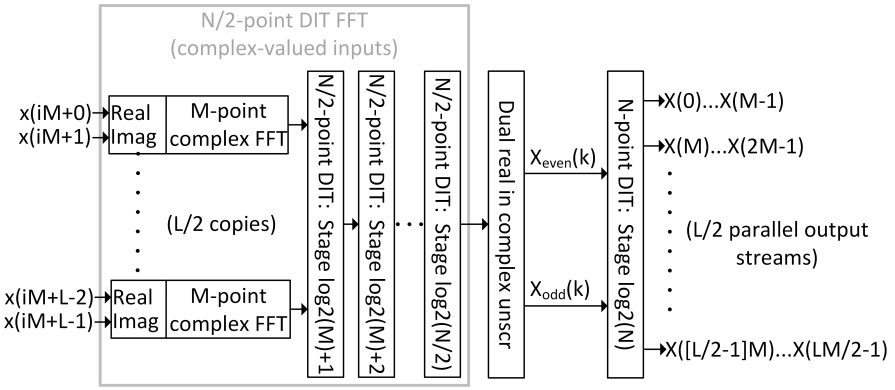


Figure 2.2: The wideband FFT architecture presented.

Computational hardware in Virtex 6 (and newer) FPGAs

FPGAs consist of a large array of reconfigurable lookup-tables (LUTs) and small memory elements (D-Flip Flops, or DFF)[2] which can be connected to each other through a configurable interconnect grid. These pieces of hardware are colloquially known as "logic", and are generally more flexible and less efficient than specialized hardware elements listed below. Programming the FPGA consists of describing the configuration of these lookup tables, memory elements and interconnect system. In order to accelerate common, computationally intensive tasks, Xilinx also provides "block hardware elements" which, while more restricted than regular logic, operate at higher clock rates, using fewer resources. The DSP48E1[3] is the Virtex 6 block arithmetic unit (henceforth DSP). Amongst many other things, the DSP48E1 can compute $P = (A + D) \times B + C$, where A, B, C, and D are 25, 18, 48, and 25-bit two's complement numbers, respectively. Internal cascade carry chains allow for the transmission of P to an adjacent DSP, saving FPGA logic and routing resources. The Virtex 6 block memory element is a 18-bit wide by 1024 address deep random-access block memory (often abbreviated BRAM in general, or RAM18 for the

[2] http://www.xilinx.com/support/documentation/data_sheets/ds150.pdf
[3] http://www.xilinx.com/support/documentation/user_guides/ug369.pdf

Xilinx-specific component). Bit widths in the multipliers and memories imply a natural data-path word width of 18 bits, which is used both in CASPER and this work. Aggressive and efficient use of these fixed hardware blocks is crucial to efficient implementation of signal processing algorithms.

2.3 Optimization of Complex Multiply Operations

A complex valued multiply, computing

$$c_r + jc_i = (a_r + ja_i)(b_r + jb_i) : c_r = a_r b_r - a_i b_i;\ c_i = a_r b_i + a_i b_r, \tag{2.4}$$

is an important component in complex-valued filter-banks such as FFTs. As of this writing, the current CASPER Library[4] uses one DSP element each for the four multiplies and implements the two addition/subtracts required in FPGA logic (Figure 2.3).

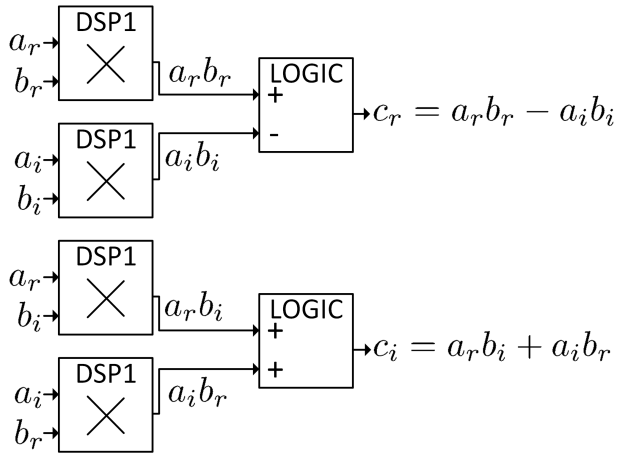


Figure 2.3: Traditional CASPER complex multiply.

A more efficient alternative, the Gauss Complex Multiplication (GCM) algorithm (Knuth, 1988), can compute a complex multiplication using three multiplications and five addition/subtractions (as opposed to the standard four multiplications and two additions): a 3-DSP complex multiply (Figure 2.4) can be performed using this algorithm. With careful instantiation of DSP48E1 elements, these modules can be made with very few pipeline registers outside of DSP elements. Normally, this would require excellent placement of the DSP48E1 elements, but a clever trick can ensure excellent co-location of these DSPs. Xilinx tools know that DSPs, which are connected via their PCIN/PCOUT ports, must be placed adjacent to each other

[4] https://github.com/ska-sa/mlib_devel/commit/0b949cbe2081cb4484cd3ccbb425810b82984566

for the ports to work. This co-location requirement is still applied, even if the DSP is not configured to use the information provided by this connection. Using this constraint, the three DSPs are always adjacent.

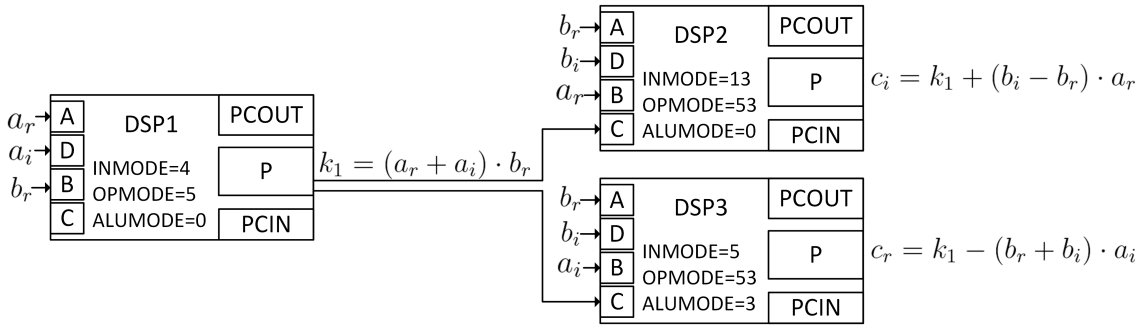


Figure 2.4: Improved complex multiply, with manual DSP configuration annotations.

2.4 FFT Optimizations

The Discrete Fourier Transform (DFT) is a filter-bank which efficiently describes a finite-length discrete signal as a function of frequency. The Fast Fourier Transform (FFT) is a divide-and-conquer implementation of a DFT which (in many cases) uses a "butterfly" as the fundamental building-block. In the implementations discussed here, these FFTs consist of repeated instantiation of these butterflies, with only coefficients and input sources changing. This paper will discuss radix-2 and radix-4 FFT implementations, which restricts the algorithms to 2^n and 4^n-point FFTs, respectively. Each FFT uses a different "butterfly", called here the "radix-2 butterfly" and "radix-4 butterfly".

Existing radix 2 FFT butterflies

CASPER uses a Decimation in Time (DIT) radix-2 FFT (and butterfly) which accepts two complex-valued inputs (A, B) and one complex-valued coefficient (ω), and computes $y = a + b\omega$; $z = a - b\omega$. The CASPER implementation uses the naive complex-multiply, followed by four real additions (each implemented in a DSP).

An extremely efficient butterfly (S. Gowda, private communication) can be implemented in five DSPs and a small amount of logic (Figure 2.5). Because DSPs are capable of performing Single Instruction Multiple Data (SIMD) operations(1x48-bit add, 2x24-bit adds or 4x12-bit adds), the final two additions are performed by the same DSP by limiting the adder input to no more than 24 bits.

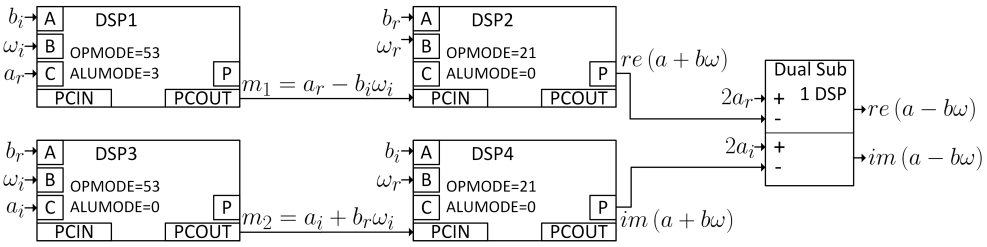


Figure 2.5: Improved 5-DSP butterfly, with manual DSP configuration annotations. The Dual Sub module performs two subtraction operations in a single DSP.

Optimized radix 4 Butterfly

A radix-4 butterfly see W. Smith, 1995b has many of the same useful properties of a radix-2 butterfly, many operations can be removed by factoring out trivial factors of $j, -1, -j$:

$$\begin{aligned} w &= a\omega^0 + b\omega^1 + c\omega^2 + d\omega^3 \\ x &= a\omega^0 - jb\omega^1 - c\omega^2 + jd\omega^3 \\ y &= a\omega^0 - b\omega^1 + c\omega^2 - d\omega^3 \\ z &= a\omega^0 + jb\omega^1 - c\omega^2 - jd\omega^3 \end{aligned}$$

A radix-4 FFT using radix-4 butterflies is restricted to 4^n-point FFTs, limiting achievable channel resolutions (unless additional effort is applied). Because each radix-4 FFT stage is equivalent to two radix-2 stages, and each radix-4 stage requires $1/2$ as many butterflies as a radix-2 stage, the radix-4 butterfly performs the work of four radix-2 butterflies. Using an efficient complex multiply combined with the ability to perform two add/subtracts in the same DSP produces a very efficient radix-4 butterfly (Figure 2.6):

Results of this optimization can be found in Table 2.1. Coefficient storage is not included in any case. One major contributor for the decrease of logic utilization for the radix-4 case is that half as many commutators are needed, when compared to the radix-2 stages. One important limitation of using a radix-4 FFT is that it restricts the user to FFT lengths of 4^n (as opposed to 2^n in the case of Radix-2 FFTs). Although this is a minor issue in the case of the first M-point FFTs (restricting the granularity of achievable frequency resolutions), it is a considerable issue for the later, L-point FFTs: restricting the size of these puts limitations on the achievable demux factor. This is resolved by adding an optimized radix-2 FFT stage to the end of the L-point FFTs, to accommodate these non $L \neq 4^n$-point FFTs.

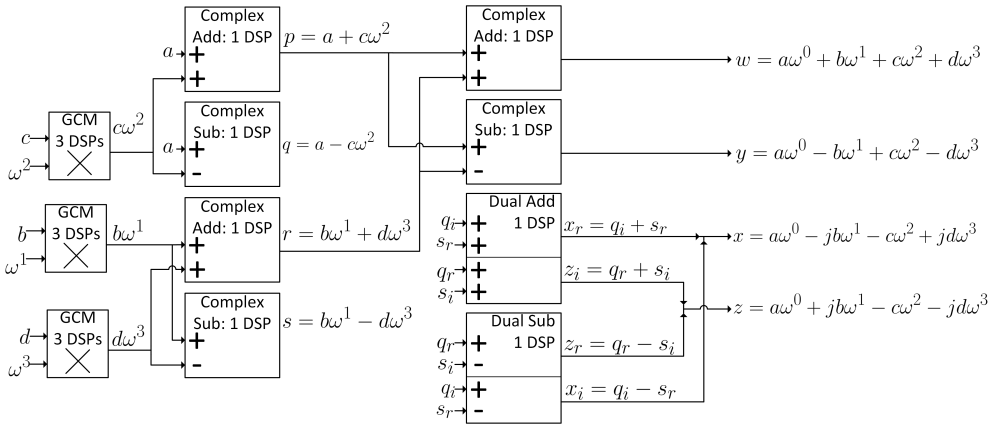


Figure 2.6: Radix-4 butterfly, implemented in 17 DSPs with minimal additional logic.

Because DSPs compute a 18x25-bit multiply, input data and coefficients for a butterfly are limited to no more than 18 bits, whereas the use of SIMD for the subsequent additions and subtractions quantizes the data internally to 24 bits. Noise analysis (discussed below) reveals that quantization at the multipliers dominates the Radix-4 butterfly's noise performance and that of the optimized Radix-4 FFT.

Module	LUT	DFF	DSP
CASPER / dual radix-2 stage 9&10	1920	3182	32
This work / dual radix-2 stage 9&10	1272	1464	20
This work / radix-4 stage 9&10	714	894	17

Table 2.1: Radix-4 FFT stage resource utilization for this work, versus other Radix-2 implementations.

Other optimizations for the FFT

Other optimizations are included in the FFT, but will only be mentioned briefly in this publication for brevity.

In the first stage of a Radix-4 FFT, all coefficients equal 1; in a radix-2 FFT, the first stage coefficients equal 1, and second stage coefficients equal {1, -1}. This allows those stages to be simplified by removing multiplications (or in the case of the radix2/stage2, replacing the multiplication with a subtraction-from-zero).

A delay-commutator FFT (He and Torkelson, 1998) such as the CASPER architecture includes delay elements between serial-FFT stages: for a radix-2 FFT, each delay is 2^{M-s} samples long (where M is the length of the serial FFT and s is the stage

number). For delays 512 cycles or shorter and certain signal bit widths (notably including the 18-bit signals popular in Xilinx FPGAs), it is possible to treat the *dual port*, 18-bit×1024-word BRAM as *two single-port*, 18-bit×512-word BRAMs. Xilinx tools do not automatically make this optimization. The same method can be used to save coefficient memory for sufficiently short coefficient ROMs. This makes it practical to justify placing more delays and coefficients in BRAM, further saving FPGA logic resources. However, because the trick only applies to short memories, the resource savings are not as impressive as those of previous techniques.

Results of FFT optimization

In order to test the resource utilization of the FFT, one of each was instantiated in a Virtex-6 design. The designs were synthesized and analyzed for resource utilization. Because the CASPER FFT is highly configurable, multiple instances were used, the most resource efficient is shown here. In the case of this work's FFT, a configuration was chosen which has successfully been compiled at 312.5 MHz on a XC6VSX475T-1 FPGA for the Starburst project (Monroe et al, in prep). Unlike previous comparisons, coefficient storage is included in these tallies. One contribution to CASPER logic utilization is unnecessary optimizations in coefficient RAM storage, which strives to reduce the total number of coefficients stored when the coefficient count was already small enough to fit in a single BRAM element, wasting logic without saving memory. As shown in Table 2.2, this work uses 1/2 to 1/4th the resources of a traditional CASPER FFT. Due to the limitations of Xilinx multipliers, intermediate FFT products are quantized to 18 and 24 bits at various processing steps. Likewise, coefficients are limited to no more than 18 bit resolution.

Module	LUT	DFF	DSP	RAM18
CASPER	38708	85565	544	142
This work	11396	16529	250	128
Fractional usage	0.294	0.193	0.459	0.901

Table 2.2: FFT (N=16384, L=16) resource utilization for traditional CASPER versus this work.

FFT timing closure

This FFT has been used in earth-observing and astronomical applications at FPGA clock frequencies of up to 375 MHz. Because of resource saving efforts, minor efforts in timing closure are needed to meet timing for high resource utilization or clock rates above 225 MHz. An overview of basic timing closure methods has been

Module	LUT	DFF	DSP	RAM18
CASPER	103923	224373	1216	139
This work	18299	30158	460	129
Fractional usage	0.176	0.134	0.378	0.928

Table 2.3: FFT (N=32768, L=32) resource utilization for traditional CASPER versus this work.

discussed in CASPER memo 50 (Monroe, 2013).

2.5 Polyphase FIR Filter Optimization

CASPER uses a Polyphase FIR as an efficient way of implementing a Weighted-Overlap-Add FFT, computing

$$p(n) = \sum_{a=0}^{T-1} c(\mod(n, N) - Na)\, x(n - Na),$$

where $x(n)$ are the input samples, N is the length of the corresponding FFT, T is the number of taps in the filter, and c is the coefficients used to define the FFT channel shapes. Often with CASPER,

$$c = w(n)\, sinc\left(\frac{\left[n - \frac{NT+1}{2}\right]}{N}\right),$$

where $w(n)$ is a window function. Implementation-wise, a polyphase FIR filter can be thought of as a bank of classical "Direct form I" FIR filter, only with delays the length of the corresponding FFT window. The number of FIR filters is equal to the demux factor for the PFB (or equivalently, the corresponding FFT). There are three ways by which the polyphase FIR filter can be optimized:

Coefficient reduction through symmetry

The coefficients of these polyphase FIR filters are nearly symmetric: the traditional CASPER filter uses a periodic window of length $NT + 1$, discarding the final sample. Shortening this window to NT results in a perfectly symmetric window. If this is done, the dual-port nature of the coefficient memories can be used to half the number of coefficients (and therefore, coefficient BRAMs) used.

Delay memory reduction through BRAM packing

Block memories are used for the long FIR filter delays. These block memories have two data-ports, each of which are 18 bits wide and 1024 addresses deep, meaning

that the optimal signal width for delays is 18-bits (for delays of 513 samples or larger), or 36-bits (for delays of 512 or fewer; as mentioned in Section 2.4, this is done by utilizing the two memory ports to treat the 1024*18 bit memory as a dual 512*18 bit memory). Packing multiple delays into the same BRAM can therefore save considerable BRAM resources (Xilinx does not infer this optimization automatically). Typical (> 3 GSPS) ADCs have no more than 8-bits, so this represents an approximately 77% delay BRAM savings (delays 512 or less), or 55% savings (delays 513 or more).

DSP/logic reduction through adder packing

As mentioned before, a DSP can perform both a multiply and an add. In the native CASPER tree, all multiplies are performed, followed by an adder-tree to sum the products. Although this would normally require T DSPs (for multiplications), plus T-1 DSPs (for additions) per FIR filter, Xilinx typically automatically packs $ceil(\frac{T-1}{2})$ into the multiplier DSPs, for a total of $2T - 1 - ceil\left(\frac{T-1}{2}\right)$ DSPs per filter. By explicitly instantiating DSP primitives, all adds can be performed in the multiplier DSPs, thus requiring a total of T DSPs per filter. Also saved is the routing logic required to connect the datapath between DSPs: the PCIN/PCOUT ports are wired directly between DSPs. Figure 2.7 illustrates the difference between these architectures.

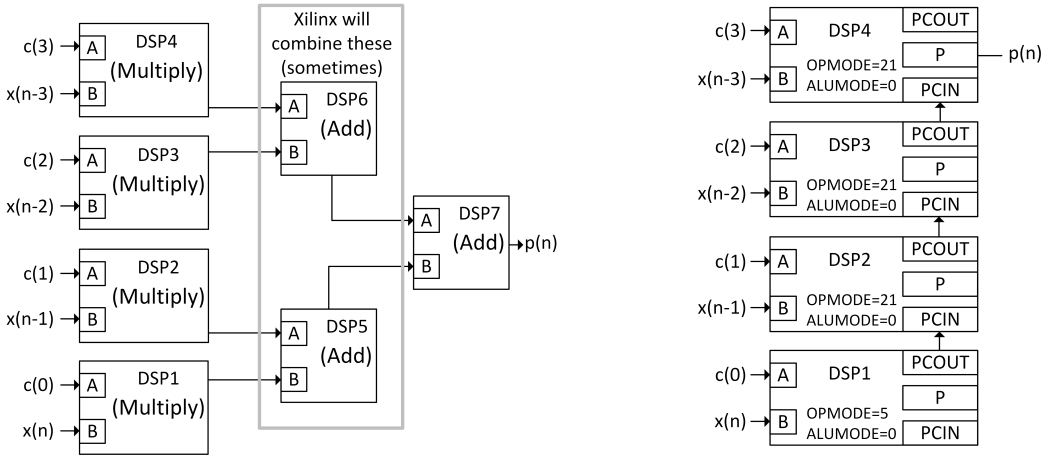


Figure 2.7: Left: original CASPER design; Right: Optimized design from this work.

Results of Polyphase-FIR optimization

A Polyphase-FIR module was instantiated in both the CASPER library and using this work (Table 2.4). Reasonable parameters were chosen for a ~5 GSPS transform.

Multiple instances of the CASPER instance were generated with various parameters: the most efficient outcome is presented here. This work's Polyphase FIR was generated using the same parameters as a design which has been compiled at 382 MHz. The parameters (N=16384, demux=16) result in a delay length of 1024; an implementation with a smaller N would have saved even more BRAM resources. It is unknown what the maximum clock rate of the CASPER implementation is, except that it is likely far below 382 MHz. Due to the size of Xilinx 18×25 bit multipliers, data input and coefficient resolution is limited to 18 and 25 bits, respectively. Output resolution is not limited.

Module	LUT	DFF	DSP	RAM18
CASPER	1906	5555	176	120
This work	1140	2494	128	57
Fractional usage	0.598	0.448	0.727	0.445

Table 2.4: Polyphase-FIR (N=16384, L=16, T=8) resource utilization for traditional CASPER versus this work.

2.6 Noise Performance

The noise performance of large DSP algorithms is generally limited by quantization of data and filter coefficients. All novel works presented in this paper are restricted to inputs no wider than 18-bits, due to the maximum multiplier size. In order to perform detailed performance analysis, a bit-true model for the algorithm was written in MATLAB. This model allows users to simulate the algorithm with quantization turned on/off in any permutation – a useful feature for determining the dominant cause of noise in the system, and therefore tuning the algorithm for minimum resource utilization at a given noise performance level.

The noise performance of fixed-point FFTs varies with input signal properties. As signals progress through the FFT, the largest magnitude signal generally increases. In order to avoid saturation or wrapping of signals (which would ruin the FFT output), signals are divided by a constant value (typically a factor of 2 for efficiency). However, dividing by larger constants effectively upscales the noise produced by intermediate products, which implies that the constants should be chosen to be as small as possible. Broadband signals distribute power into FFT bins roughly equally, requiring minimal scaling. Narrow-band input signals integrate output power into a small number of channels, requiring more aggressive scaling.

A synthetic input $x(n)$ was generated and passed through three versions of the

fixed-point model: one operating on only double-floating-point values, one which quantized data only at the simulated ideal ADC, and finally one which quantized data in a manner which is bit true to a typical instantiation of this algorithm. The output spectra are denoted $X_{FP}(k)$, $X_{ADC}(k)$ and $X_{QUANT}(k)$, respectively. The noise contributed by the ADC and algorithm quantization are defined as $N_{ADC}(k) = \mathbb{E}\left(|X_{FP}(k) - X_{ADC}(k)|^2\right)$ and $N_{QUANT}(k) = \mathbb{E}\left(|X_{ADC}(k) - X_{QUANT}(k)|^2\right)$. The algorithm was instantiated with parameters (N=16384, taps=8, demux=16, coeff/Data quantized to 18 bits, ADC quantized to 12 bits) and driven with shaped gaussian noise $\mathcal{N}\left(0, \frac{1}{5}\text{adc\_fullscale}\right)$. Results can be seen in Figure 2.8. Note that performance will vary with scaling schedule, and therefore input signal characteristics. That said, with these, the algorithm remains ADC noise dominated for most of the spectrum. This author typically uses the algorithm with 8-bit ADCs, where the relative noise performance is ~24 dB better.

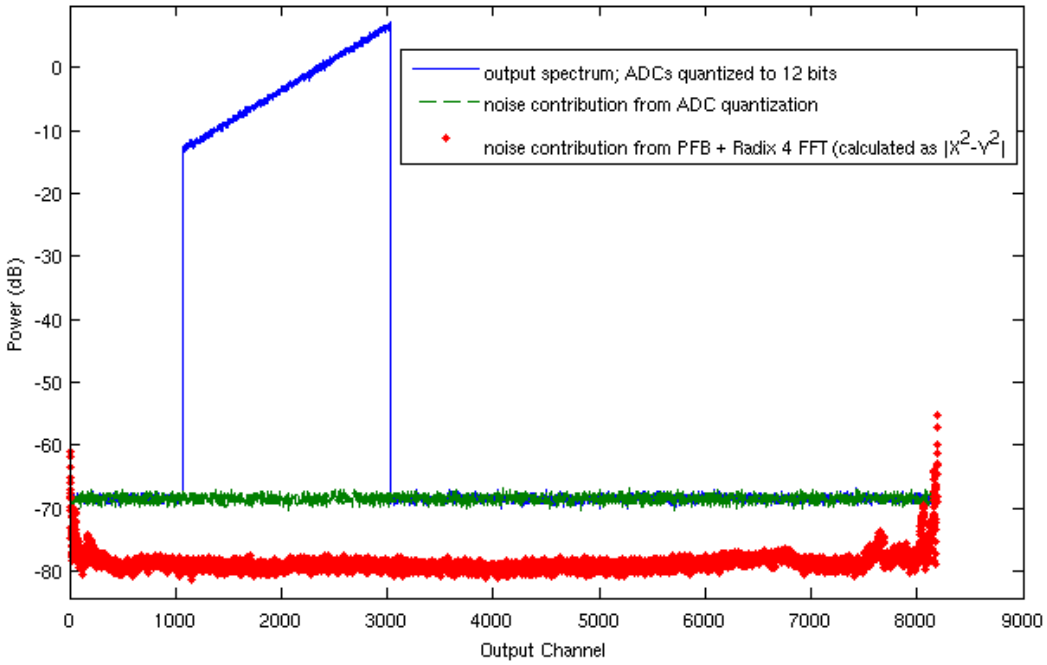


Figure 2.8: Noise analysis of optimized radix-4 FFT.

2.7 Conclusions

I have presented a very efficient implementation of commonly-used CASPER algorithms. The Polyphase FIR design is strictly superior to the stock CASPER implementation (in both a resource and statistical sense), and may be used as a drop-in replacement in all designs which currently use CASPER's Polyphase FIR

algorithm. Although it is harder to analyze the noise performance of the FFT, it is believed to be mathematically equivalent to CASPER implementation, and utilizes far fewer resources than the original. Given its advantages, the FFT described here is a more suitable choice for CASPER-based designs with an instantaneous FFT bandwidth of 800 MHz or more. It is worth noting that in cases where the spectral resolution is not available (due to the 4^K restriction of the initial, serial, FFT), this algorithm at the next higher resolution step is often still more efficient than using CASPER at the original resolution. These algorithms are freely available on Github[5].

2.8 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. AST-1311098.

References

A. Oppenheim, R. Schafer (2009). *Discrete-time signal processing*. 3rd ed. Prentice-Hall, pp. 623, 723, 746.

He, Shousheng and M. Torkelson (1998). "Designing pipeline FFT processor for OFDM (de)modulation". In: *1998 URSI International Symposium on Signals, Systems, and Electronics. Conference Proceedings (Cat. No.98EX167)*, pp. 257–262. DOI: 10.1109/ISSSE.1998.738077.

Knuth, D. (1988). *The Art of Computer Programming*. Vol. 2. Addison-Wesley, pp. 519, 706.

Lyons, R. (2010). *Understanding Digital Signal Processing*. Prentice-Hall. Chap. 10.4, 13.20.

Monroe, R. (2013). *Performance optimization for Virtex 6 CASPER designs*. https://casper.berkeley.edu/wiki/Memos.

W. Smith, J. Smith (1995a). *Handbook of Real-Time Fast Fourier Transforms*. Prentice-Hall, pp. 16–20.

– (1995b). *Handbook of Real-Time Fast Fourier Transforms*. Prentice-Hall, pp. 87, 88, 213–222.

[5] https://github.com/ryan-monroe/monroe_library

Chapter 3

THE STARBURST BACKEND

3.1 Introduction

The search for habitable planets has been a strong motivator for astronomy research in recent decades. One important metric for planetary habitability has been its location relative to the local "Goldilocks zone", where the temperature is able to support liquid water. However, liquid surface water can only occur if the planet has sufficient atmospheric pressure. NASA's MAVEN mission dramatically demonstrated that ion pickup by Coronal Mass Ejections (CMEs) contributes significantly to the loss of Mars' atmosphere due to the absence of a global B field on Mars (Jakosky et al. 2015). One logical conclusion for this is that planetary magnetic fields can drive planetary habitability and that the magnetic field strength required to successfully shield planets will be driven by the strength of these coronal mass ejections.

Flare stars, or magnetically active M dwarfs, produce energetic flares across the electromagnetic spectrum, whose energy far exceeds that of observed solar flares (Lacy et al. 1976). It is expected that these extreme flares are accompanied by extreme CMEs, eruptions of magnetized plasma from the 10^6-K stellar corona. Based on the high flare rate of active M dwarfs, it is predicted that stellar CMEs may dominate these stars' mass loss (Osten & Wolk 2015), non-radiative energy output (Drake et al. 2013), and angular momentum evolution (Aarnio et al. 2013). Around an active M dwarf, CMEs could potentially erode an entire planetary atmosphere (Khodachenko et al. 2007, Lammer et al. 2007) and/or continually deplete almost all ozone from an Earth-like atmosphere (Segura et al. 2010). However, these predictions all rely on extrapolating the observed relationship between solar flares and CMEs (e.g., Yashiro & Gopalswamy 2009) , as there are no definitive observations of stellar CMEs.

Wide bandwidth radio spectroscopy offers the potential to directly detect stellar CMEs, using a method that has long been used to study shock waves in the solar corona (Nelson & Melrose 1985). The shock front formed by fast-moving solar CMEs produces bright radio emission at low harmonics of the local plasma frequency, which is determined by the electron density. As a CME moves outwards from the Sun, it moves into lower densities and therefore the radio emission drifts to

lower frequencies. The frequency drift rate can be used to measure source speeds and study density structure in the solar corona. As discussed by Crosley et al. 2016, this method may be extended to nearby stars to directly detect stellar CMEs. To do so requires a radio facility with the sensitivity for stellar observations (cooled receivers, large collecting area) and with wide fractional bandwidth ($\nu_{\rm max}/\nu_{\rm min} \gg 1$). Wide fractional bandwidth allows tracking the motion of sources across a wide range of distances from the star, and it also enables categorization of the different types of stellar radio bursts, helpful for identifying the specific signatures of the CME-associated bursts.

Starburst engineering: challenges and motivations

The Starburst system was designed for wideband radio spectroscopy of nearby flare stars, with the goal of detecting stellar CMEs. The motivation of Starburst was to make a wideband interferometer between two 27m antennas at Owen's Valley Radio Observatory (OVRO). In the initial design phase of the Starburst analog and digital backend electronics, it became clear that the requirements for Starburst overlap significantly with other projects. Therefore, we designed this system to be a low-cost, efficient implementation which can be replicated for a variety of project specifications, with minimal effort to the end-user. Aside from Starburst itself, the system is well-suited to applications that require wide bandwidth such as the detection of Fast Radio Bursts (FRBs) or pulsars, as well as mapping of CO lines.

Recent progress in microwave receiver and feed design have delivered an ultra-wideband multi-octave capability which, if its full potential is to be realized, requires a broadband analog and digital back-end. Digital hardware such as the CASPER ROACH2s has been designed to meet these demands. CASPER additionally provides digital libraries to support FPGA code development (J. Hickish et al., 2016). We apply the libraries in (R. Monroe, 2016) and additionally present modular algorithms for channelization as well as integration of higher order statistics. The provided libraries provide a ~50 per cent reduction in resource utilization for these expensive algorithms, with a corresponding reduction in the number of ROACH2s required for most radio astronomy applications. Here we build on the libraries introduced in (R. Monroe, 2016) by introducing a new design for an analog and digital backend (the latter utilizing these same libraries) that delivers flexible, low-cost ultra-wideband spectroscopy and correlation capability with excellent resource utilization and a suite of additional diagnostic tools built in as modular features.

We applied these algorithms in this design, known as the Starburst system. Starburst was originally developed for wideband radio spectroscopy of nearby flare stars, with the goal of detecting stellar CMEs (Lacy, Moffett, and Evans, 1976; Jakosky et al., 2015; Khodachenko et al., 2007; Lammer et al., 2007). However, it has additionally been designed as an inherently flexible standalone system that can be replicated and reused with minimal design effort to the end-user. By designing with flexibility and testability in mind from the ground up, this firmware allows circumvention of the substantial design and testing typically required in designing systems for radio astronomy applications, especially to meet timing under the extremely demanding clocking requirements that wide bandwidth signal processing requires. This allows variations of the system to be deployed for additional science cases that particularly benefit from this ultra-wideband capability such as solar spectroscopy, the detection of Fast Radio Bursts (FRBs) and pulsars, as well as mapping of CO lines.

We present here the design and initial commissioning results for the Starburst back-end, including results of in-lab testing, and a demonstration of successful cross-correlation on an astronomical calibrator.

3.2 System Requirements

The top-level system requirements for the original Starburst project were driven by both the science demands and cost. Building on the recent front-end upgrade as a part of the Expanded Owens Valley Solar Array (EOVSA) project (Gary, 2016; Gary et al., 2014) provided the opportunity to use the two 27 m antennas at OVRO with the Starburst system for a number of hours each day. EOVSA was retrofitting each of the antennas with two interchangeable cooled receivers using wideband dual-polarization feeds.

To match the bandwidth of the EOVSA 1-6 GHz feed (Akgiray et al., 2013), we designed the Starburst system to have an instantaneous bandwidth of 5 GHz. At the time, the ADC1x5000-8 (Homin Jiang and J. Hickish, 2013), an 8-bit, 5 Gsps was the fastest well tested (Jiang et al., 2012) ADC board for the ROACH2. This restricted us to sidebands no wider than the Nyquist rate of 2.5 GHz. To minimize cost and maintain flexibility, we chose to use a sideband separating downconversion design, which makes it possible to change bands by changing the LO frequency and does not require separate filters for each band.

One secondary concern was system harmonics: harmonic structure is one of the identifying characteristics of stellar CME-associated radio bursts, but previous ex-

perience with solar observations has shown that instrumental effects can mimic this harmonic structure (see footnote 3 of Bastian et al. 1998). Therefore the Starburst system was designed with the specification that spurious harmonic structure should be weaker than -30 dB. Finally, the Starburst system was designed to facilitate RFI flagging by the user with minimal loss of adjacent portions of the spectrum, by offering flexible high time and frequency resolution and an option to calculate the spectral kurtosis correlation product (Nita and Gary, 2010).

Insurmountable structural issues with the 27 m dishes were discovered during preliminary commissioning of the Starburst system, ending the project. As a consequence, demonstration of the system is limited to lab tests, detections of geostationary satellites, and detections of Cygnus A, a bright astronomical source.

3.3 System Overview

The key parameters, a top level diagram and the Bill of Materials of Starburst can be found in Table 3.1, Figure 3.1 and (Villadsen and R. Monroe, 2015), respectively. Astronomical signals are received by the two 27 m antennas at OVRO. Two selectable (cooled) receivers are available: one covering 1-6 GHz and the other covering 3-18 GHz (Akgiray et al. 2012). Switching between these at the prime focus allows us to achieve the objective 1-18 GHz selectable bandwidth. After filtering and amplification by the front-end electronics, the signals are transmitted by optical fiber to a central building for processing. Here, the signal in a selectable 5 GHz band in the range 1-18 GHz is down-converted with a local oscillator (LO) in the middle of the band to an intermediate frequency (IF) from DC to 2.5 GHz. Two mixers are used, with the LO in quadrature between the two. The corresponding in-phase (I) and quadrature (Q) IFs can be combined after digitization to separate the upper and lower sidebands of the LO, recovering the full 5-GHz band (Jones, 2010). The signals are then sampled and processed in an FX engine implemented in four CASPER ROACH2 FPGA boards (Glenn Jones, 2013). Careful optimization of algorithms allowed for the high utilization of these ROACH2 boards – this task would typically be performed in ∼six ROACH2 boards due to FPGA utilization concerns. The Starburst digital system can be used to cross-correlate four inputs with an instantaneous bandwidth of 5 GHz: four single polarization inputs or two full-polarization inputs.

Parameter	Value
Polarizations Per Antenna	2
Baseline Length	400 m
System Temperature	Est. ~50 K
Instantaneous Bandwidth	5 GHz
Frequency Range	1-18 GHz
Frequency resolution	305 KHz
Selectable bands	{0.9-5.9; 5-10; 9-14; 13-18} GHz
Dump Rate	7 ms – 10 s

Table 3.1: Top-level system parameters.

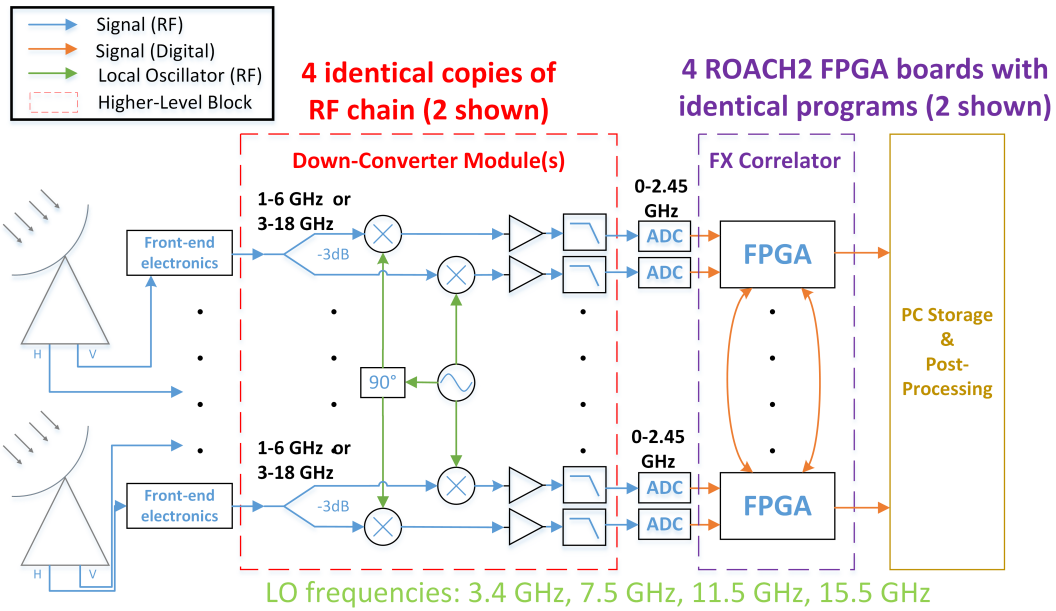


Figure 3.1: Top-level block diagram.

3.4 Analog Electronics

The RF electronics chain is displayed in Figure 3.2. All of the components used in the downconverters are commercial off-the-shelf in order to minimize cost and are listed in the BOM. Here, a 10 MHz reference tone is used as a reference to a 100 MHz Phase-Locked Dielectric Resonance Oscillator (PLDRO). The LOs, at frequencies (3.4, 7.5, 11.5 and 15.5 GHz), as well as the 2.5 GHz ADC clock, are then phase-locked to this 100 MHz signal. One LO is selected using an absorptive switch, amplified, and split using a quadrature hybrid and two four-way power dividers into four copies each of sine and cosine LO. Another switch allows for the selection between a common reference signal and the astronomical signal. The chosen input signal is then split and mixed by the sine and cosine LO signals,

creating the downconverted in-phase (I) and quadrature (Q) products, respectively. Each signal is then amplified and passed through a band-defining filter. A coupler provides a small amount of power to the power detectors as a diagnostic tool. The switches and RF power detectors, as well as temperature and voltage monitors, are all managed by LabJacks. Each of the two antennas was assigned a Labjack, which controlled the reference/astronomical switch and supply/RF power monitoring for the analog electronics associated with that antenna. A third Labjack was used to control the LO switch, as well as power supply monitoring for that system. Custom daughterboards were designed to support this task: they were identical for all three Labjacks, meaning that some circuitry was unused on each board. Further information, including prices at the time of purchase and datasheets, can be found on the Bill of Materials (Villadsen and R. Monroe, 2015).

Finally, all four "I" signals and all four "Q" samples are sampled at 5 GSPS using Casper's ADC1x5000-8 (Homin Jiang and J. Hickish, 2013) (E2V EV8AQ160). Further processing is performed in the FPGAs.

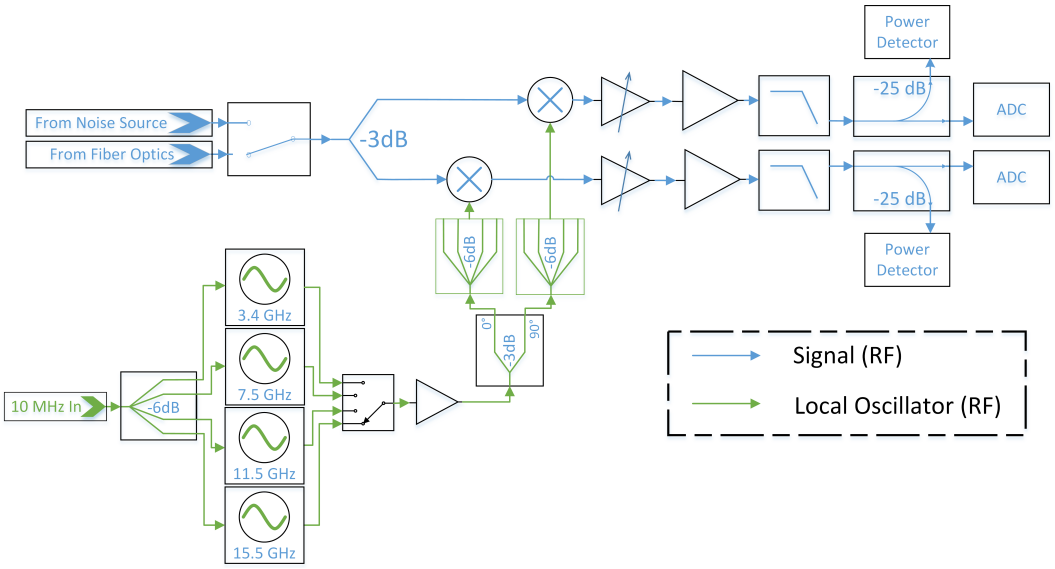


Figure 3.2: Downconversion block diagram, as described in Section 3.4.

3.5 FPGA Signal Processing

Four ROACH2s, each containing two ADC1x5000-8 boards as well as a Virtex 6 FPGA act as the FX correlator. Figure 3.3 contains a block diagram of the FPGA's firmware.

Sampling

ADC1x5000-8 boards have ~ 6.5 Effective Number of Bits (ENOB) and bandwidth to approximately 2.2 GHz: reasonable for our application, which has a maximum baseband frequency of roughly 2.45 GHz. Each of the on-board EV8AQ160 parts (Semiconductors, 2014) has four 1.25 GSPS cores which are time-interleaved to achieve a 5 GSPS sample rate. In order to use the ADCs, the SERDES communication links between FPGA and ADC must be phase-aligned to account for routing and component delays. This is performed by configuring the ADC to generate a test signal, which is compared with the signal received by the FPGA. The on-FPGA PLL clock's phase offset is then swept until the best error-free PLL phase is found. Code and a more complete description of the process can be found at (Primiani, 2017).

Time-domain interleaving of ADC cores can result in frequency-dependent gain and offset mismatches between the individual cores, which result in spurious signals in the frequency domain (Razavi, 2012). In order to minimize these effects, the built-in gain and offset settings available in each core are adjusted to remove the electronic mismatches. First, the system is driven with an input which is characteristic of that during normal operations: typically, broad-band noise. Then, raw ADC samples are captured for each offset setting of each core, with the gains configured to their default setting: the offset setting for each core which minimizes the mean value of the raw ADC samples is used. Finally, the process is repeated for the gain settings, setting the RMS of raw samples to the mean RMS of all cores when they are at their default gain setting and optimal offset setting. Code to perform this task is included alongside all other starburst-related code (R. Monroe and Villadsen, 2018).

F-engine

Each ROACH2 performs the F-engine tasks for a single polarization of one antenna. On each ROACH2, two ADC1x5000-8 boards sample the input signal at 8 bits/5 GSPS. For each signal, a 4-tap Polyphase pre-filter and 2^{14}-point FFT comprise a Polyphase Filter-Bank (PFB). The algorithms used in (Monroe, 2016) were used due to the tight resource availability in this design. I/Q separation is performed per (Thompson, Moran, and Swenson, 2001; Jones, 2009). Lab measurements of the frequency response of each signal chain allow for the upper-sideband (USB) and lower-sideband (LSB) signals to be digitally recovered, in spite of a gain mismatch between the signal-chains.

One concern for interferometrs is visibility decorrelation, as the dot product of

the baseline and the source direction changes over the course of an integration. Assuming a linear phase sweep over the course of an integration, the relative SNR for a monochromatic source can be found as $SNR_{relative} = \frac{||\int_{-\Delta\phi/2}^{\Delta\phi/2} e^{j\theta} d\theta||}{\Delta\phi}$. Using a fiducial metric of no more than 2.5 per cent loss in SNR, this can be easily solved to find $\Delta\phi = 0.77$ radians. The change in projected baseline length due to the Earths rotation can be found as $\Delta\phi = \frac{2\pi B}{\lambda} \sin(\omega_e * t_{max})$, or $t_{max} = \frac{\sin^{-1}\left[\frac{\Delta\phi * \lambda}{2\pi * B}\right]}{\omega_e}$, where the rate of the rotation of the Earth, $\omega_e = 7.27e-05$ rad/s. For our instrument with a baseline $B = 418$ and $\lambda_{min} = 1.7$ cm, this implies a maximum integration length of 67 ms. By removing this time-dependent geometric effect ("Fringe Stopping"), this decorrelation can be removed in real-time. This is done by applying a phase ramp to the spectrum: $\theta(n,k) = 2\pi k^n/_N$ radians, where k is the sample delay, n is the frequency channel, and N is the total number of points in the FFT (in this case 16 384). Finally, the data are re-quantized from the (18 bit real / 18 bit imag) values which are internally used, to (4 bit real / 4 bit imag) values which are transmitted over 10 Gigabit Ethernet (10 GbE) to the X-engine (cross-correlator).

X-engine

In order to save hardware, the same four ROACH2 boards are simultaneously used as the X-engine. Each board performs the cross-correlations for $\frac{1}{4}$th of the spectrum ($\frac{1}{2}$ of one sideband). Signals are routed from the F-engine segment of a ROACH2 board to the X-engine segment of the appropriate ROACH2 via 10GbE. In the X-engine, cross-correlation products are computed. Although auto-correlations could also be produced using the 4 bit resolution data, they are instead computed on the ROACH2 board which hosts that F-engine (Fourier Transform), where full 18 bit resolution data can be used. For the same reason, the fourth moment of each polarization (used in computing the kurtosis statistic) and the ADC sample histograms are likewise computed on the ROACH2 which hosts that signal's F-engine. Optionally, auto-correlations and cross-correlations may be accumulated in frequency by summing 2^n channels, where n is an integer. If $n \neq 0$, the fourth moment and histogram statistics are invalidated due to limitations in the digital design.

Controls and Communication

Interfacing appears to be one of the areas containing the largest variety across spectroscopy projects. Depending on the application, projects require a tremendous variety of integration lengths, integration triggers (especially in projects which use a temperature reference), meta-data, and output formats. The Starburst firmware

seeks to address all of these varied requirements in a single package.

All integration resets, coarse delay changes, and sideband separation coefficient updates may be optionally synchronized to the 1-Pulse-Per-Second (1PPS) signal. Integration resets may be optionally delayed from that 1PPS signal to allow for transients from coarse delay changes to pass through before integration begins again: the system latency is approximately 8 FFT windows long, or roughly 26.2us, making this a common choice for this setting.

Data products are sent via One Gigabit Ethernet (1 GbE) to a host computer: the maximum data-rate is limited by the bandwidth on this 1 GbE link. In order to maximize dump-rate for non-Starburst applications, subsets of the data products may be omitted: for a given output product (such as auto-correlations or fourth moments), frequencies channels within the range [1024*n, 1024*n+1023; n=0...7], for each subband, may be collectively transmitted or omitted.

The Starburst backend was designed with flexibility in mind: the goal is for the firmware to be usable out-of-the-box in most wideband, small-N spectrometers and correlators. Sideband separation can be easily disabled; integration length in time and frequency is run-time configurable; all packets include detailed metadata about system state and timing; accumulations can be optionally reset via software or 1PPS, with configurable reset delays. Furthermore, General Purpose Input/Output (GPIO) pins are connected to either software control or various system parameters, at the user's choice. By default, these pins are used to broadcast the system state over the on-board LEDs. The GPIO input pins can also be used to reset integrations or be read out by the controlling computer. Finally, a bit-true model of much of the firmware is available, implemented in MATLAB. The compiled firmware, as well as all source code and documentation, is available at (R. Monroe and Villadsen, 2018). Many algorithms from (R. Monroe, 2016) are used: the PFB and FFT are instantiated directly, while the X-engine DSP is modified from the same library's algorithms. All other DSP in the system is implemented specifically for the Starburst project. Interface and infrastructure code (both software and firmware) utilize substantial contributions from the CASPER libraries.

The Starburst firmware's output rate is limited by the 1GbE output bandwidth limit of $1e9$ b/s. The firmware and link have been tested to approximately 60 per cent of that capacity (the limit of the host capture software at this time). Bandwidth requirements for each output product sum linearly and are described in Table 3.2.

Table 3.2: Bandwidth requirements for Starburst Firmware output products (assuming full polarization), given an integration dump rate R (H), number of baselines B, and channel count C, the latter of which is driven by desired channel resolution. Channel selection is constrained to the 1024-channel blocks mentioned above, while fourth moments, histograms and test patterns are only valid for full resolution spectra. Starburst used $R = 10$, $B = 6$, $C = 16\,384$, resulting in a total datarate of 88.5 Mbps, or about $1/12$th of the maximum bandwidth.

Product	Bandwidth Requirement
Cross-correlation (excludes autocorrelations)	$48BCR$
Cross-correlation between local ADCs	$96CR$
Fourth moment	$96CR$
ADC histograms	$96 \cdot 2048 \cdot R$
Debug test pattern	$48 \cdot 1024 \cdot R$

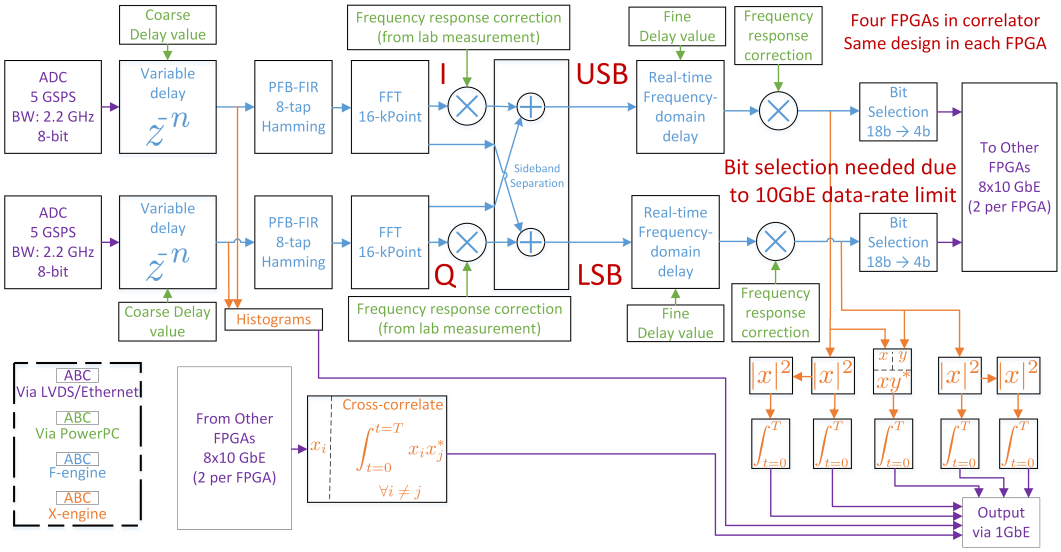


Figure 3.3: FPGA firmware block diagram.

3.6 Laboratory Testing

A number of tests were performed to demonstrate this system: bandpass shape, PFB channel shape, instantaneous sideband separation performance, stability of sideband separation (not shown in this paper), and sideband separation performance with under-sampled calibration points. These tests are detailed below.

Because broadband digital sideband separation had not been demonstrated as of the design of this system, we believed that this was the most likely point of failure and therefore tested it aggressively. In order to confirm successful digital sideband separation, the calibration coefficients were calculated using a tone synthesizer configured to sweep in frequency – the same system was then used to test the per-

formance of that separation. One concern with this analysis is that because the calibration and subsequent separation are being performed at the same frequencies (as opposed to the neighborhood around the calibration point), the test results could be optimistic. By under-sampling the calibration process, we put limits on the severity of this optimism. Finally, in order to characterize RFI mitigation performance in the presence of strong RFI, the shapes of individual channels were measured, showing that the channel-to-channel RFI mitigation meets our specifications.

To test the digital sideband separation described above, an HMCT2100 programmable sine generator was connected to the system input. This generator was swept in frequency, allowing for a measurement of the relative response between the I and Q channels of the system. After calculating and applying I/Q separation coefficients, the tone sweep process was repeated. This allowed for measurements of the system's sideband separation performance, which is shown in (Figure 3.4, right). Through the same procedure, we measured the response of the "I" channel of one of our RF+digital backend modules, which is shown in (Figure 3.4, left).

It is important to note that the I/Q separation measurements made in the above plot are made at the same frequencies as calibration, which makes them something of a "best case" in terms of performance. In order to gain a measure of the deterioration of performance against frequency, the system was configured to perform sideband separation calibration on data from every 71 channels (21.6 MHz resolution), but measure the separation performance at full (305 kHz) resolution. The results of that test are shown in (Figure 3.6/left, center). The reduction in separation performance away from the measurement points is most likely due to a mis-synchronization between the ADCs serving the I and Q channels of this system. Due to limitations in the ADC-FPGA communications infrastructure, the ADCs may be misaligned in time by even numbers of samples (up to 8 in either direction). Inspection of the cross-correlation between ADCs shows that the misalignment was 6 samples in this dataset (Figure **3.6**/right), which implies a maximum sideband separation ratio of approximately 29.6 dB at a point between the two measurement locations, assuming that all system variation is caused by this phase ramp. In this dataset, cubic spline interpolation was performed to bridge the gaps between measurement points – this removes most of the effects of the delay mismatch, enabling our better than 29.6 dB maximum separation. As a consequence, it is important that future researchers who intend to perform digital sideband separation ensure good ADC timing matches (through some combination of good digital design, coarse delay

correction and possibly repeated reboots of the digital system until the ADCs are synchronized correctly), as well as good I/Q chain matches in general, to minimize the cost of calibration. An imperfect coarse delay match imposes a phase ramp on the calibration coefficients. This phase ramp results in frequencies within each channel requiring a different coefficient, whereas the system is only capable of assigning one coefficient for every four channels, as a performance/speed tradeoff. While it is likely that this phase ramp was a limiting factor for sideband separation performance with the Starburst system, we had more than achieved our separation leakage specification and therefore accepted the system as it is.

Using the same tone-sweep technique, the frequency response of individual channels from the F-engine were tested. Good rejection channel-to-channel will provide strong rejection of narrowband RFI. Additionally, a well-designed system will have individual channels which bear a similar or identical frequency response. The results of this test can be seen in Figure 3.5. This test demonstrates a strong 45-50 dB channel-to-channel rejection, along with consistent frequency response between channels. There is some variation in the total power response between channels which is undesirable (Figure 3.5;Right): simulation indicates that this is not caused by the digital algorithms, but it is unclear which of the Tone Generator, DCM and ADCs are contributing to this small ~0.03 dB variation.

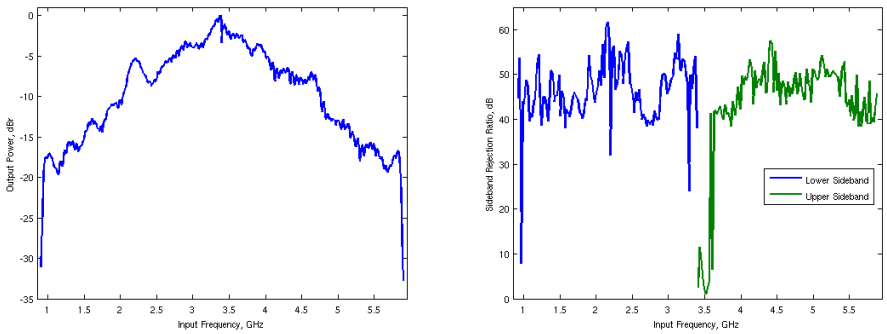


Figure 3.4: **Left:** Starburst frequency response measured for the combined RF and digital backend. **Right:** Achieved I/Q separation as a function of frequency. The spurs at {0.965, 2.20,3.28, 3.608} GHz are due to unreliable frequency switching behavior in the input source: these are removed in software through outlier detection and coffficient spline fitting for normal operation; the poor separation from 3.41-3.95 GHz is unexplained.

If the project was not halted due to the aforementioned antenna issues, we would have performed additional tests, especially of system stability and ADC performance.

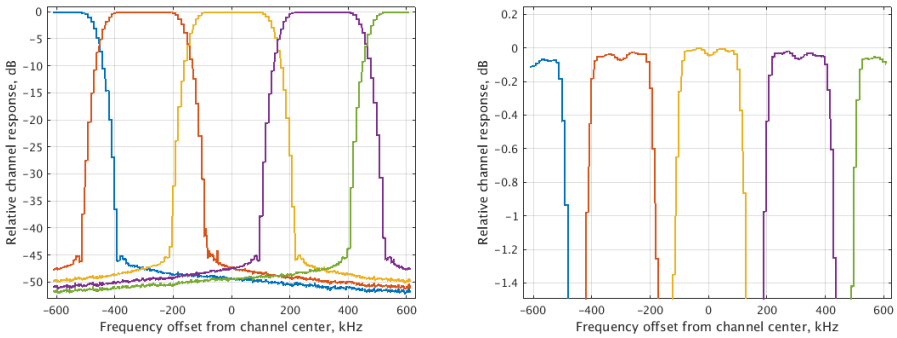


Figure 3.5: **Left:** Starburst instrument channel shapes, as measured through down-conversion and digital systems. **Right:** Same plot with vertical zoom. Staircase in frequency is due to the frequency resolution of the test source.

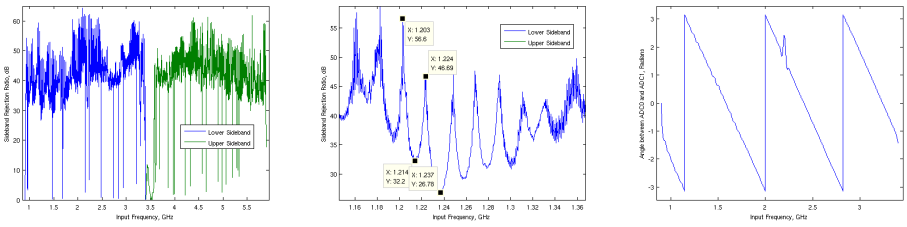


Figure 3.6: **Left:** Sideband separation ratio, after calibrating at every 71th channel instead of at full resolution. **Center:** Zoomed version of Left. Note the regular pattern in performance variation, which is discussed above. **Right:** Phase angle of the cross-correlation between I and Q ADC samples. The phase ramp implies a 6-sample offset between I and Q.

Although preliminary versions of these tests were performed and showed promising results, this author does not believe they were sufficiently controlled to be included in an instrument publication such as this. All data is available upon request to the author (R. Monroe).

3.7 First Light Results

A test observation of Cygnus A (Figure 3.7) was performed to demonstrate the operation of the system. This test observation used a single-polarization cross-correlation between one 27 m antenna and a 2 m antenna from the OVSA. There were many important challenges in this test observation because of the limited passband of the small Solar dishes. An ideal result would include visibilities with magnitude approximately equal to the product of the spectrum of Cygnus A with the frequency dependent gain of the antennas. In the frequencies which have sufficient gain, a "healthy" spectrum would show coherent phases with a linear phase ramp.

The spectrum produced does see this region of elevated gain and coherent phase, but the visibilities are much noisier than we expect. It is likely that the low-frequency gain dropoff is caused by a 2.1 GHz high-pass filter on the 2 m OVSA antennas. The cause for the high phase noise from 3.5-4 GHz and 4.5-6 GHz is unknown but was not investigated due to the aforementioned antenna faults, which were discovered at approximately the same time. While the project was discontinued at this point due to the structural instabilities in the antennas, this observation demonstrated the ability of the Starburst back end to successfully detect fringes from an astronomical radio source.

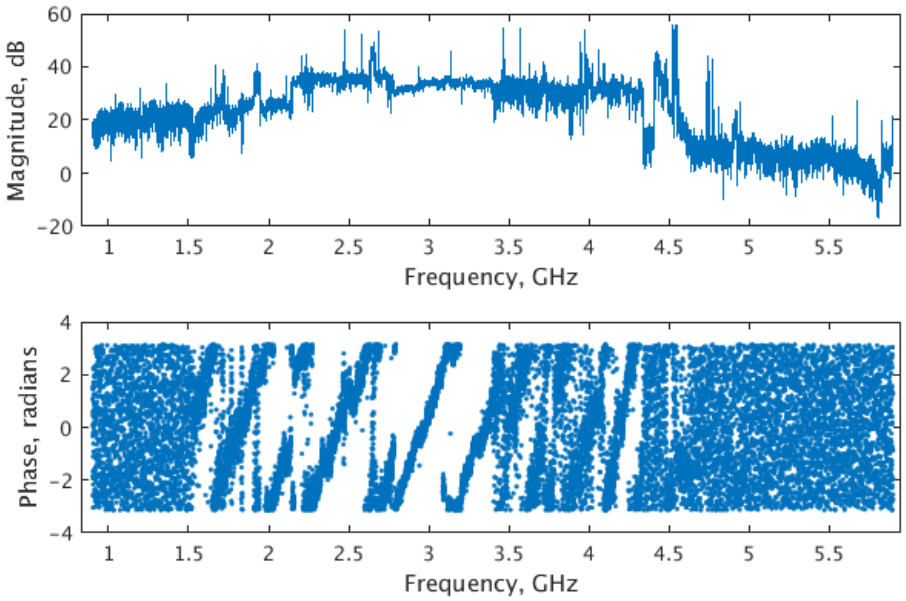


Figure 3.7: Uncalibrated visibilities of CygA, as observed with the Starburst system at 17:29 PST, October 9 2016. Integration length was one second.

3.8 Relevance to Other Projects

Starburst is highly applicable for implementation as a 2 antenna correlator; the same infrastructure can be applied to high time/frequency resolution spectrometer and interferometer projects, such as earth-observing science, transient monitoring systems, pulsar backends and Carbon Monoxide mapping (COMAP) (Li et al., 2016). For instance, dumping auto-correlations at a frequency resolution of 305 kHz is possible at a rate of approximately 1 ms. Current backends could find the Starburst digital backend to be a low-cost and easy upgrade to their existing system for wider instantaneous bandwidth.

3.9 Conclusions

In the beginning of this paper, we referenced the need for a versatile, wideband digital backend, designed to be used in a wide variety of scientific instrumentation applications. We have presented a powerful digital answer to this challenge, coupled with a simple, low-cost I/Q downconversion scheme. Both downconverter and correlator were designed for the Starburst Project but with the flexibility to be used in a wide variety of radio-astronomy and earth observing science applications. This system has been tested in the lab and at OVRO. Full design documents are available at (R. Monroe and Villadsen, 2018). The author invites users to email him with any issues.

Acknowledgements

Starburst is funded by the NSF (AST-1311098). We would like to thank the Xilinx University Programme for their donation of FPGAs to this project.

References

Akgiray, A. et al. (2013). "Circular Quadruple-Ridged Flared Horn Achieving Near-Constant Beamwidth Over Multioctave Bandwidth: Design and Measurements". In: *IEEE Transactions on Antennas and Propagation* 61, pp. 1099–1108. DOI: 10.1109/TAP.2012.2229953.

Gary, D. E. (2016). "Early Observations with the Expanded Owens Valley Solar Array". In: *AAS/Solar Physics Division Meeting*. Vol. 47. AAS/Solar Physics Division Meeting, p. 301.01.

Gary, D. E. et al. (2014). "The Expanded Owens Valley Solar Array (EOVSA)". In: *American Astronomical Society Meeting Abstracts #224*. Vol. 224. American Astronomical Society Meeting Abstracts, p. 123.60.

Glenn Jones, Wesley New (2013). *ROACH-2 Revision 2*. https://casper.berkeley.edu/wiki/ROACH-2_Revision_2.

Hickish, J. et al. (2016). "A Decade of Developing Radio-Astronomy Instrumentation using CASPER Open-Source Technology". In: *Journal of Astronomical Instrumentation* 5, 1641001-12, pp. 1641001–12. DOI: 10.1142/S2251171716410014. arXiv: 1611.01826 [astro-ph.IM].

Homin Jiang, Jonathan Weintroub and Jack Hickish (2013). *ADC1x5000-8*. https://casper.berkeley.edu/wiki/ADC1x5000-8.

Jakosky, B. M. et al. (2015). "MAVEN observations of the response of Mars to an interplanetary coronal mass ejection". In: *Science* 350, 0210, p. 0210. DOI: 10.1126/science.aad0210.

Jiang, Homin et al. (2012). "Digitizing The Yuan Tseh Lee Array for Microwave Background Anisotropy by 5Gsps ADC boards". In: *2012 19th IEEE International Conference on Electronics, Circuits, and Systems (ICECS 2012)*. DOI: 10.1109/icecs.2012.6463741.

Jones, Glenn (2009). "Instrumentation for wide bandwidth radio astronomy". PhD thesis. California Institute of Technology.

Khodachenko, M. L. et al. (2007). "Coronal Mass Ejection (CME) Activity of Low Mass M Stars as An Important Factor for The Habitability of Terrestrial Exoplanets. I. CME Impact on Expected Magnetospheres of Earth-Like Exoplanets in Close-In Habitable Zones". In: *Astrobiology* 7, pp. 167–184. DOI: 10.1089/ast.2006.0127.

Lacy, C. H., T. J. Moffett, and D. S. Evans (1976). "UV Ceti stars - Statistical analysis of observational data". In: *The Astrophysical Journal Supplement Series* 30, pp. 85–96. DOI: 10.1086/190358.

Lammer, H. et al. (2007). "Coronal Mass Ejection (CME) Activity of Low Mass M Stars as An Important Factor for The Habitability of Terrestrial Exoplanets. II. CME-Induced Ion Pick Up of Earth-like Exoplanets in Close-In Habitable Zones". In: *Astrobiology* 7, pp. 185–207. DOI: 10.1089/ast.2006.0128.

Li, T. Y. et al. (2016). "Connecting CO Intensity Mapping to Molecular Gas and Star Formation in the Epoch of Galaxy Assembly". In: *The Astrophysical Journal* 817, 169, p. 169. DOI: 10.3847/0004-637X/817/2/169. arXiv: 1503.08833.

Monroe, R. (2016). "Improving the Performance and Resource Utilization of the CASPER FFT and Polyphase Filterbank". In: *Journal of Astronomical Instrumentation* 5, 1641002-262, pp. 1641002–262. DOI: 10.1142/S2251171716410026.

Monroe, Ryan and Jackie Villadsen (2018). *Github - Starburst backend design*. https://github.com/ryan-monroe/starburst_backend_design.

Nita, G. M. and D. E. Gary (2010). "The generalized spectral kurtosis estimator". In: *Monthly Notices of the Royal Astronomical Society* 406, pp. L60–L64. DOI: 10.1111/j.1745-3933.2010.00882.x. arXiv: 1005.4371 [astro-ph.IM].

Primiani, Rurik A. (2017). *Github - sma-wideband adc tests*. https://github.com/sma-wideband/adc_tests.

Razavi, Behzad (2012). "Problem of timing mismatch in interleaved ADCs". In: *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*. DOI: 10.1109/cicc.2012.6330655.

Semiconductors, e2v (2014). *EV8AQ160 QUAD ADC*. https://casper.berkeley.edu/wiki/images/1/19/Ev8aq160.pdf.

Thompson, Richard, James Moran, and George Swenson (2001). *Interferometry and Synthesis in Radio Astronomy*. Wiley-VCH, pp. 181–183. ISBN: 0471254924. URL: https://www.amazon.com/Interferometry-Synthesis-Astronomy-Richard-Thompson/dp/0471254924?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=0471254924.

Villadsen, Jackie and Ryan Monroe (2015). *Starburst master BOM*. https://tinyurl.com/y8je8vkv.

Chapter 4

RFI MITIGATION AT NANOSECOND TIMESCALES

4.1 Introduction

Radio Frequency Interference (RFI) is a major challenge for all modern radio observatories. Although some frequency bands are reserved for radioastronomy, these are rarely broad enough to support the tremendous bandwidth able to be processed by broadband digital back-ends. Many modern radio observatories contend with RFI produced by sources such as artificial satellites (A. Offringa et al., 2015; Ellingson, Bunton, and Bell, 2000), ground communications, and arcing power lines (Crane, 2010). RFI is a serious impediment to radio-astronomy: LOFAR flags a few percent of it's data due to RFI (A. R. Offringa et al., 2010), and the production of terrestrial and space-borne RFI has been showing a upward trend. This paper will discuss the analysis of two similar RFI sources which were later discovered to be produced by arcing power lines.

4.2 Current RFI mitigation techniques

The techniques by which RFI is mitigated are as diverse as the family of signals themselves. Some methods rely on the assumption that astronomical signals are smooth in time or space: sharp deviations may indicate RFI (Fridman and Baan, 2001). Many of these–such as the popular AOFlagger (A. Offringa et al., 2012) operate on post-correlation visibilities. Likewise, impulsive event detection can be performed on raw ADC samples (K. D. Buch et al., 2017) – often by flagging the FFT window associated with detected impulsive RFI. Techniques have also been developed which use higher-order statistical moments of the signals such as Spectral Kurtosis to detect low duty cycle RFI events (Nita et al., 2007; Misra, Roo, and Ruf, 2012). If the interference is mostly produced by a small number of stationary powerful sources, cross-correlation between reference antenna(s) and the remainder of the array allows for good removal of the signal (Mitchell and Robertson, 2005; Briggs and Kesteven, 2000; Barnbaum and Bradley, 1998). Reference antenna techniques are only useful if the number of RFI sources is small and it is economical to produce an antenna with sufficient gain to point at the RFI source–challenging at long wavelengths. In situations where this is not possible, subspace-based techniques often work under the assumption that RFI is caused by small number of point sources

which are stationary on timescales of an integration. Projecting out the dominant singular vectors produces a spatial null in the direction of the RFI (Raza, Boonstra, and Veen, 2002), but typically also projects out some astronomical data–special care must be taken when this is an issue.

4.3 The Owen's Valley Long Wavelength Array

The OVRO-LWA is an all-sky viewing telescope located in Owen's Valley, California, which cross-correlates 256 dipoles across the band 30-88 MHz, with a frequency resolution of 24 KHz. Digital processing is implemented by the popular FX correlator architecture. The F-engine (an input-wise frequency transform, often implemented using a Polyphase-FFT) is implemented using 16 ROACH2 FPGAs boards (Glenn Jones, 2013) each equipped with HMCAD1510 ADCs (Dexter and al., 2014) to sample the 512 inputs at 196.608 MSPS (Nyquist bandwidth of 98.304 MHz) with an 8-bit resolution (32 inputs per ROACH2). The F-engine performs time-domain delay correction, channelizes the spectrum using a 8192-point Polyphase Filterbank, performs digital bandpass leveling, and quantizes the resulting 18-bit real/imaginary number to 4 bits real/imag for transmission via 10 Gigabit Ethernet (10GbE) and subsequent processing and imaging. An image of the daytime radio sky as seen by the OVRO-LWA can be seen in Figure 4.1: RFI seen at the horizon is powerful enough to produce artifacts (in the form of side-lobes) which affect the entire image. This RFI has strong time-frequency structure (Figure 4.2) which appears to correlate with a day-night cycle.

4.4 RFI analysis

This section depicts the process of characterizing the RFI that is clearly prominent in OVRO-LWA images, particularly near the horizon. Without correction, it a dominant noise source for LWA's science-grade data, but at first the specific nature of the source (power line RFI) was unclear. We characterized the source and produced a filtering mechanism to blank appropriate temporal epochs on-chip. A fast-capture mode was developed for OVRO-LWA, allowing for raw capture of the complex-valued pre-correlation frequency-domain voltage signal. The FPGAs produce 4-bit real/4-bit imag samples, at the FFT channel resolution of 24 kHz. First, we analyzed our ability to detect these events in the antenna's auto-correlation data by summing the power of all samples across antenna, polarization and frequency. This approach is attractive because it does not require knowing an accurate calibration in real-time. Later we repeated this analysis, instead beam-forming the array in

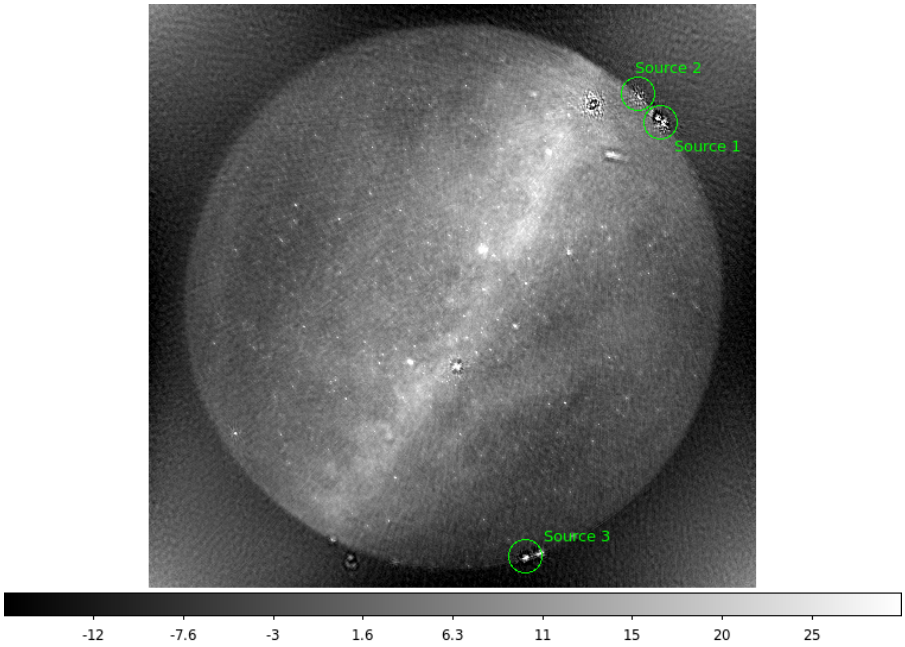


Figure 4.1: The night sky, as seen from OVRO. Circle indicates the horizon RFI in question.

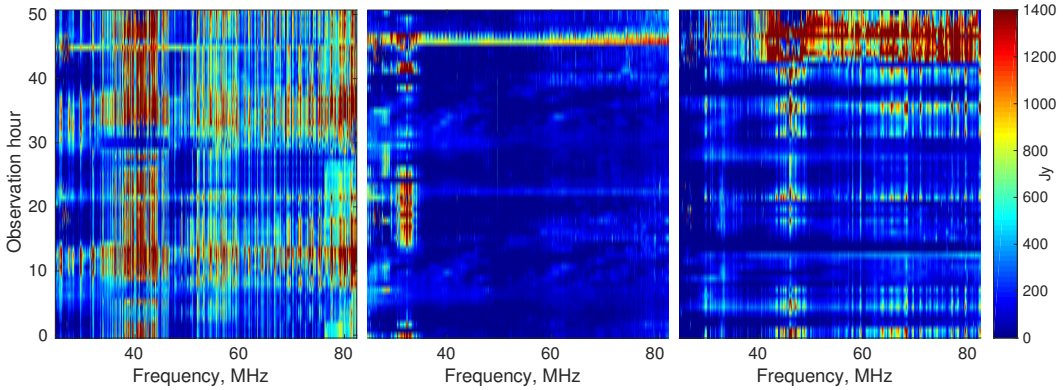


Figure 4.2: From left to right: the power spectrum of sources 1, 2, and 3. Obtained by making an image at each time/frequency point and observing the peak power near the direction of the stationary RFI source.

the direction of the RFI–beamforming suppresses the rest of the sky for a more sensitive measurement. Although most of the analysis displayed below has also been performed on the auto-correlation data, we found (Figure 4.13) that the beamformed detections were necessary to mitigate the RFI to an acceptable level. For this reason as well as brevity, this work will mostly display the beamformed effort.

The RFI's characteristics are clear from a spectrogram (Figure 4.8) of the data at a $41\mu s$ time-resolution. The powerful, regular spikes are clearly non-astronomical and

indicative of RFI–however, their impulsive pulse profile is producing a broadband signature. Operating on 2.62 MHz of bandwidth (a convenient number based on the instrumental architecture), we summed across frequency and used a simple detector (described below in Section 4.5) to extract the pulses (Figure 4.4).

In assist subsequent analysis, a gradient-descent fit on frequency and phase was performed, finding a fundamental frequency of 59.99 Hz. The fit is very consistent throughout the 25s window. Performing an FFT on the power of the summed, beamformed samples show that the 60 Hz tone (and its harmonics) are extremely narrow-band over the 25s. (Figure 4.5)

To analyze the regularity of event timing, the detection threshold was adjusted such that only very clear events would be chosen. Adjacent events were de-duplicated. The true time of each remaining event was set to be the weighted average of the power at that time instance along with the two adjacent time samples. The resulting event gap histogram can be seen in (Figure 4.6). Curiously, although no events occur closer than 375us from each other for all ~30,000 pulses, there is considerable timing jitter in the exact arrival of impulsive events. A clear preferred interval such as that would have been evidence for a software-driven controller causing the pulses, which is not seen here.

Using the sinusoidal fit applied above, the pulse pattern as a function of phase can be analyzed. Figure 4.7 shows a 2-D histogram of all pulses and their powers. Also displayed is the average power of all pulses detected at each given phase. The pulse power does not appear to change with phase on the 60 Hz wave: carrier appears to merely correlate with the presence (or lack thereof) of RFI events.

There is evidence for another group of pulses, which seem to be co-located and collectively of lower power. By searching for weaker pulses (and excluding the strongest of pulses), we can selectively detect these pulses and fit a second sinusoid to them (Figure 4.8). Notably, the second sinusoid is 0.66π out of phase with the first–approximately what would be expected for interference sourced by three-phase power.

This analysis was repeated on another, less powerful RFI source to the northwest of the instrument (“Source B”). This source was found to be extremely similar to the initial source presented in this document. Special care was taken to confirm that the sources were distinct–beamforming in the direction of the original source (which has since been resolved) produces no detection. Similar analysis to that presented

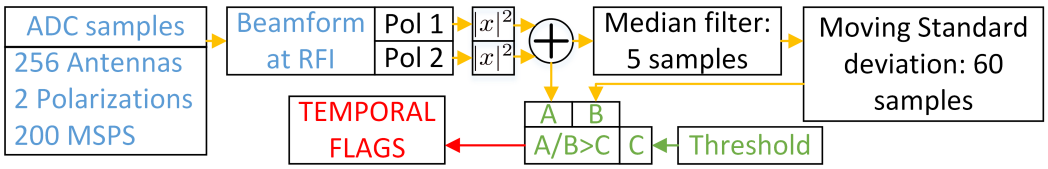


Figure 4.3: Top level block diagram of pulsed event detection algorithm.

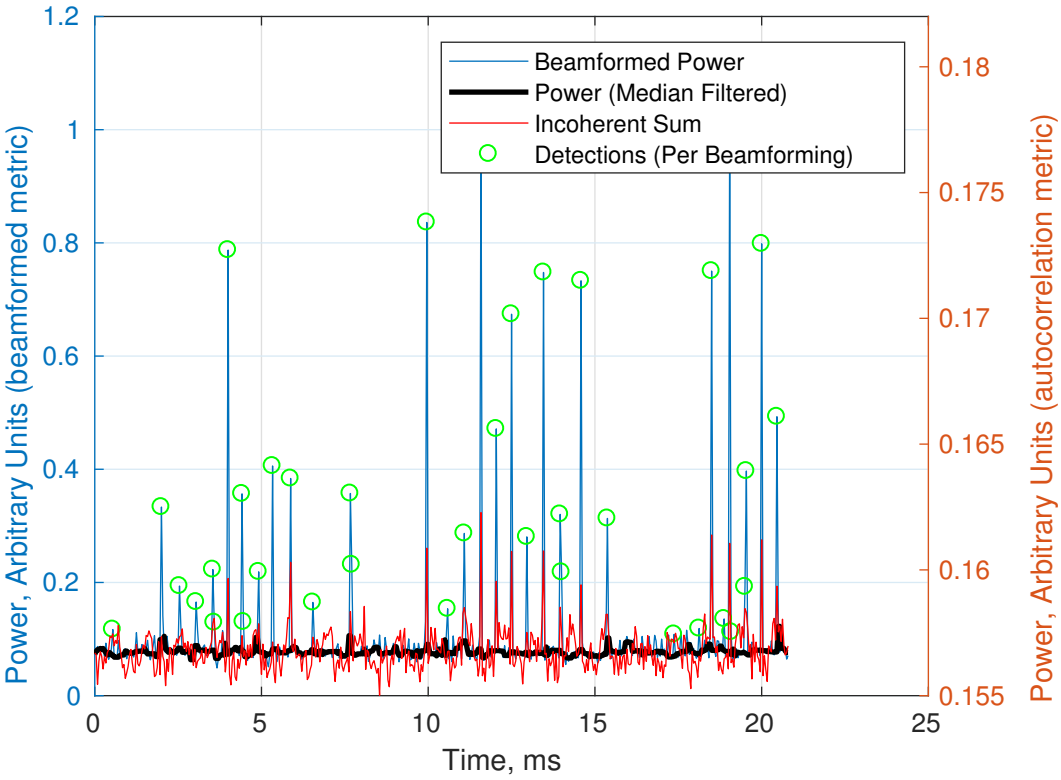


Figure 4.4: Beamformed power detections of RFI events. Blue traces represent received power beam-formed in the direction of the RFI source. Applying a median filter (black) permits a robust estimate of present astronomical power. Green circles represent detected impulsive events, and red traces represent the same statistic but with only the use of autocorrelation information. While some events are still detectable with autocorrelations, the performance is greatly reduced.

above can be found in Figures 4.9, 4.10, 4.11, 4.12).

4.5 Power line RFI mitigation

Power line RFI at low frequencies presents a unique challenge. Being pulsed, the RFI is broadband. Although powerful pulsed signals are typically a strong target for Kurtosis-based detectors (Roo, Misra, and Ruf, 2007), this RFI is very regular, showing up in every single integration. This combination of broadband and

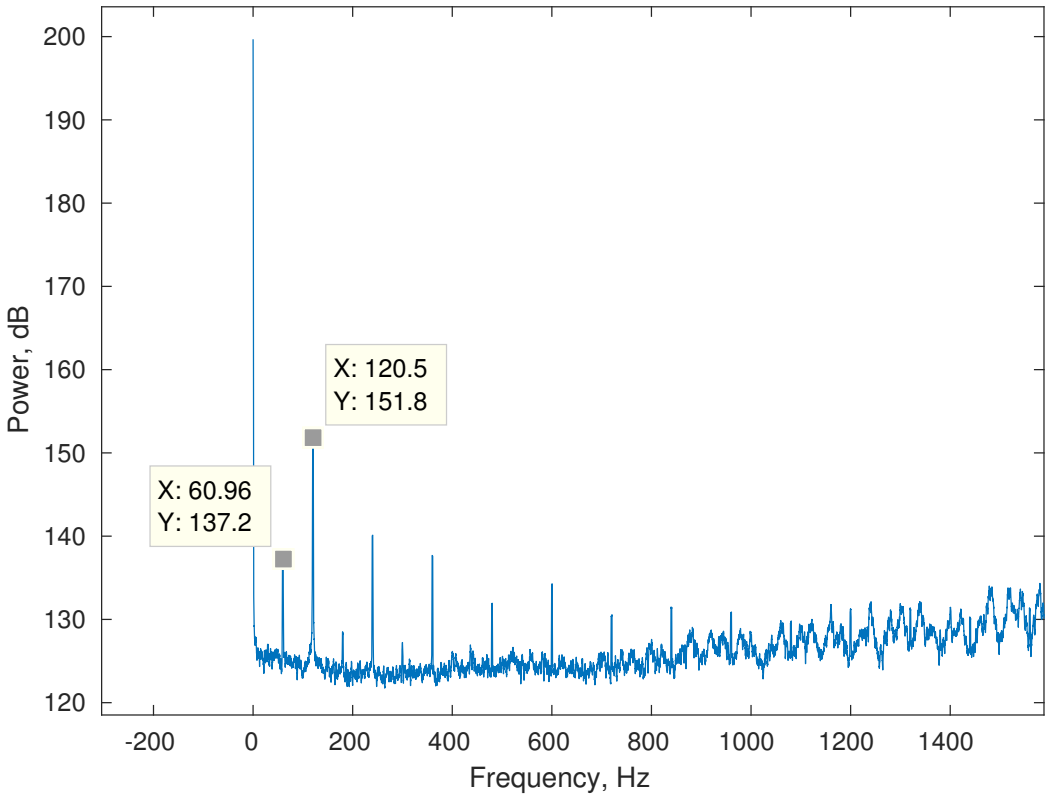


Figure 4.5: Power spectral density of the beamformed power. Higher frequency components are likely due to the gap between pulses. Power is heavily concentrated in 60 Hz and its harmonics.

high duty cycle (on 13s integration timescales) makes integration-resolution and frequency domain RFI filtering impractical: flagging must occur on much faster timescales. The low operating frequency of the array implies that any reference antenna with sufficient gain to target an RFI source would be impractically large, excluding that technique. Finally, subspace-based techniques were not preferred in fear of filtering astronomical signal as well. Fortunately, the powerful time-domain pulses imply that a simple power detector which operates on the F-engine output is sufficient, provided that it operates at a fast enough time resolution. Note that a similar detector could also be designed to operate on raw ADC samples, and may perform somewhat better on account of not being limited to 4bit real/imag signals. However, doing design for GPUs is much more convenient than FPGAs, driving the architecture described here.

First, the signals from the 256 antennas are beamformed in the direction of the RFI, as computed from images. Since the astronomical power and instrumental

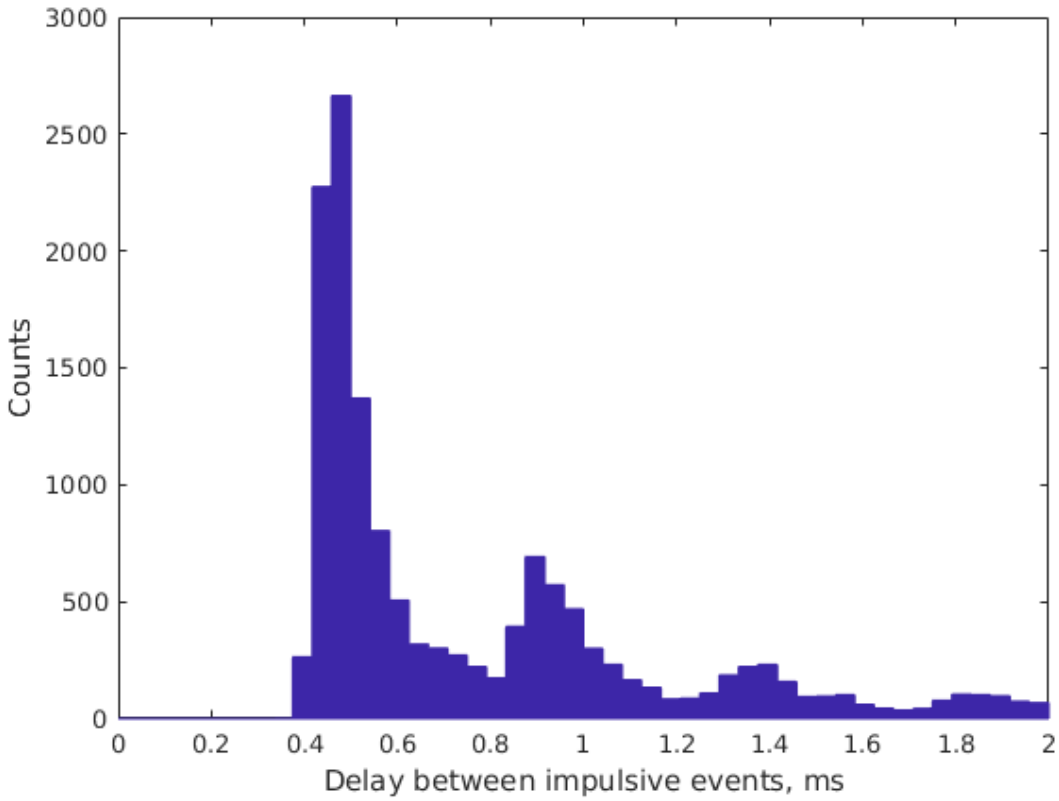


Figure 4.6: Histogram of the gap between pulse events. Only very rarely do events contain a gap longer than ~ 1.5 ms.

noise are roughly white across the 2.6 MHz band which the analysis occupied, the power in this beamformed signal is simply summed across frequency. This gives us a scalar value $P(t)$ representing the total power within this 2.6 MHz band, at a $(24\ \text{KHz})^{-1} \approx 41.6\mu s$ time resolution.

Afterwards, a simple power threshold suffices to detect the RFI events. Empirically, the strong RFI pulses occupy approximately 7% of time instances, while the weaker grouping is detected in another 4% of instances. Weaker pulses resulting in less detection sensitivity likely contributes to the lower detection rate for the weaker events.

This detector worked because the properties of the astronomical signal did not change substantially during the 25s observing window. An alternative detector was also tested which would detect these pulses in the presence of a time-variable background. Here, $M(t) = \text{medfilt}(P(t), 5)$ produces a 5-sample running median-filter on the data. This mitigates the effects of the RFI pulses, but also degrades the astronomical signal. Next, $N(t) = \sqrt{\text{movvar}(M(t))}$ estimates the noise amplitude. Finally, a

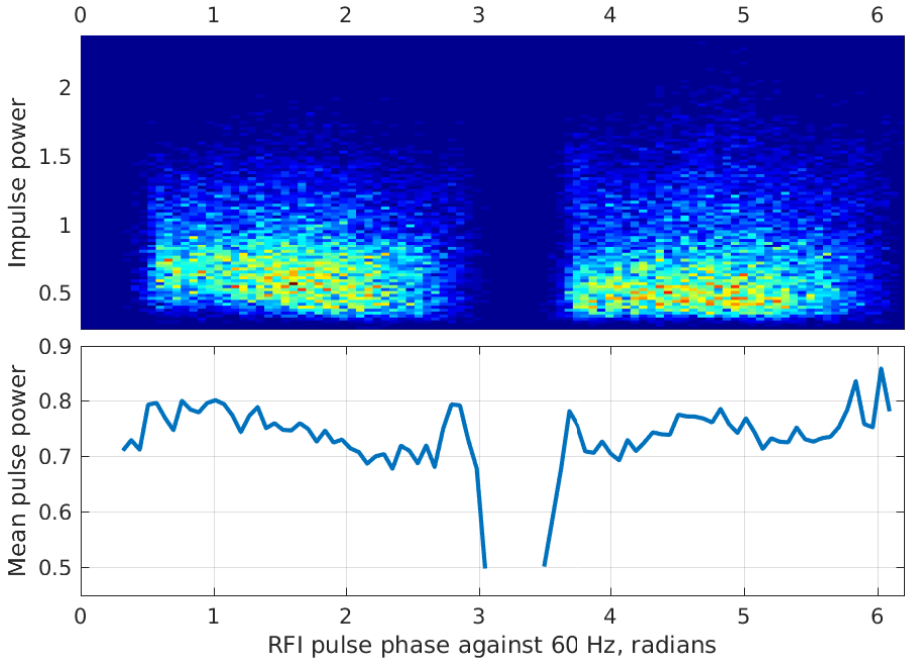


Figure 4.7: Top: Histogram of pulses against phase of a 60 Hz sinusoid. Bottom: Average power of detected pulses. The gaps at $\sim$ 0 radians and 3 $\sim$ 3.8 radians represents a region in the 60 hz mains power cycle where impulsive RFI events are extremely rare. On the other hand, pulse power does not appear to be a function of 60 hz phase.

detection statistic can be computed as $D(t) = \frac{P(t)-M(t)}{N(t)}$, and RFI is detected for all t where $t > \gamma$, a detection threshold. For a Gaussian background, this statistic will have consistent false-positive detection rate, regardless of background astronomical power variability. In hindsight, this technique bears some resemblance to that of (K. Buch, 2013), applied to beamformed signals instead of individual antenna autocorrelations.

To demonstrate this method, 25s of raw data at 2.6 MHz bandwidth was collected at the OVRO-LWA. The detection technique listed above was used, and offending data were flagged and not included in the subsequent software cross-correlation. Figure 4.13 shows the results: a reduction of residual power from the RFI source by a factor of approximately 15.9. This residual estimate assumes no power contribution from the sky in that direction–an assumption which artificially reduces the estimation of algorithmic performance.

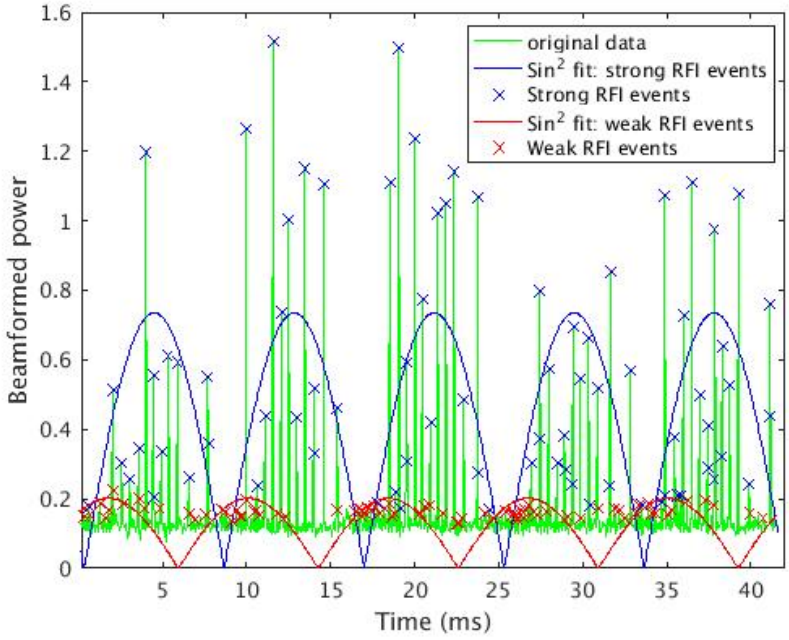


Figure 4.8: RFI detections, with an overlaid ~60 hz sine wave fit to event detections. The blue and red markers represent what is likely different phases of a 3-phase mains power signal.

4.6 Conclusions

This work presents the analysis of two instances of power-line RFI using the analysis of high time-resolution data, as produced by an interferometer's F-engine. A powerful technique which mitigates most of this interference was presented as well. This technique works on both calibrated and uncalibrated instruments, but calibration (and therefore, beamforming) produces much more impressive results. These techniques can be applied in software to those low-frequency FX correlators which have the hardware ability to produce multiple beams, such as LOFAR, GMRT, LWA, and CHIME. The full 352-antenna OVRO-LWA, will implement this in GPU, with a few beams to target troublesome RFI sources. All code used to produce these plots can be found at (Monroe, 2018).

References

A. Offringa et al. (2012). "A morphological algorithm for improving radio-frequency interference detection". In: *Astronomy and Astrophysics* 539, A95. DOI: 10.1051/0004-6361/201118497. URL: https://doi.org/10.1051/0004-6361/201118497.

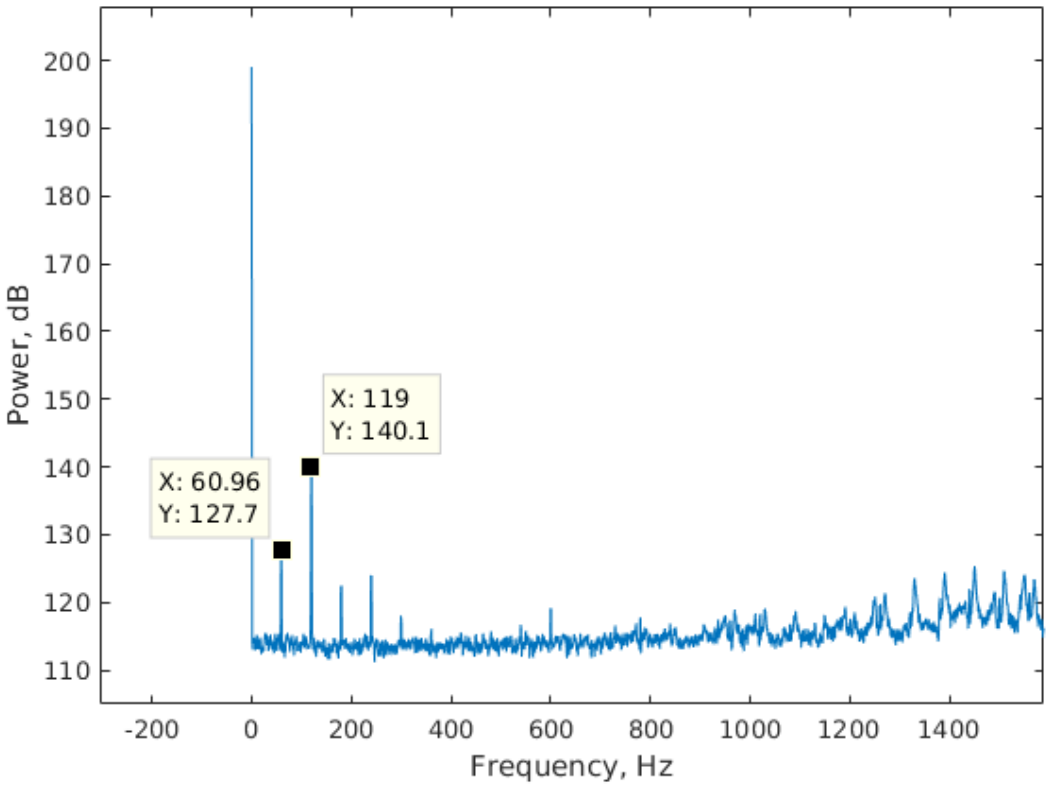


Figure 4.9: Power spectral density of the beamformed power for Source B. Higher frequency components are likely due to the gap between pulses. Power is heavily concentrated in 60 Hz and its harmonics.

Barnbaum, C. and R. F. Bradley (1998). "A New Approach to Interference Excision in Radio Astronomy: Real-Time Adaptive Cancellation". In: *The Astronomical Journal* 116, pp. 2598–2614. DOI: 10.1086/300604.

Briggs, Frank H. and M. J. Kesteven (2000). "Removing Radio Interference from Contaminated Astronomical Spectra Using an Independent Reference Signal and Closure Relations". In:

Buch, K. D. et al. (2017). "Implementing and Characterizing Real-time Broadband RFI Excision for the GMRT Wideband Backend". In: *ArXiv e-prints*. arXiv: 1710.08576 [astro-ph.IM].

Buch, Kaushal (2013). *Impulsive RFI Excision: CASPER Library Block*. https://casper.berkeley.edu/wiki/Impulsive_RFI_Excision:_CASPER_Library_Block.

Crane, Patrick (2010). *Radio-Frequency Interference (RFI) From Extra-High Voltage (EHV) Transmission Lines*. http://www.faculty.ece.vt.edu/swe/lwa/memo/lwa0168.pdf.

Dexter, Matt and et al. (2014). *ADC16x250-8 RJ45*. https://casper.berkeley.edu/wiki/ADC16x250-8.

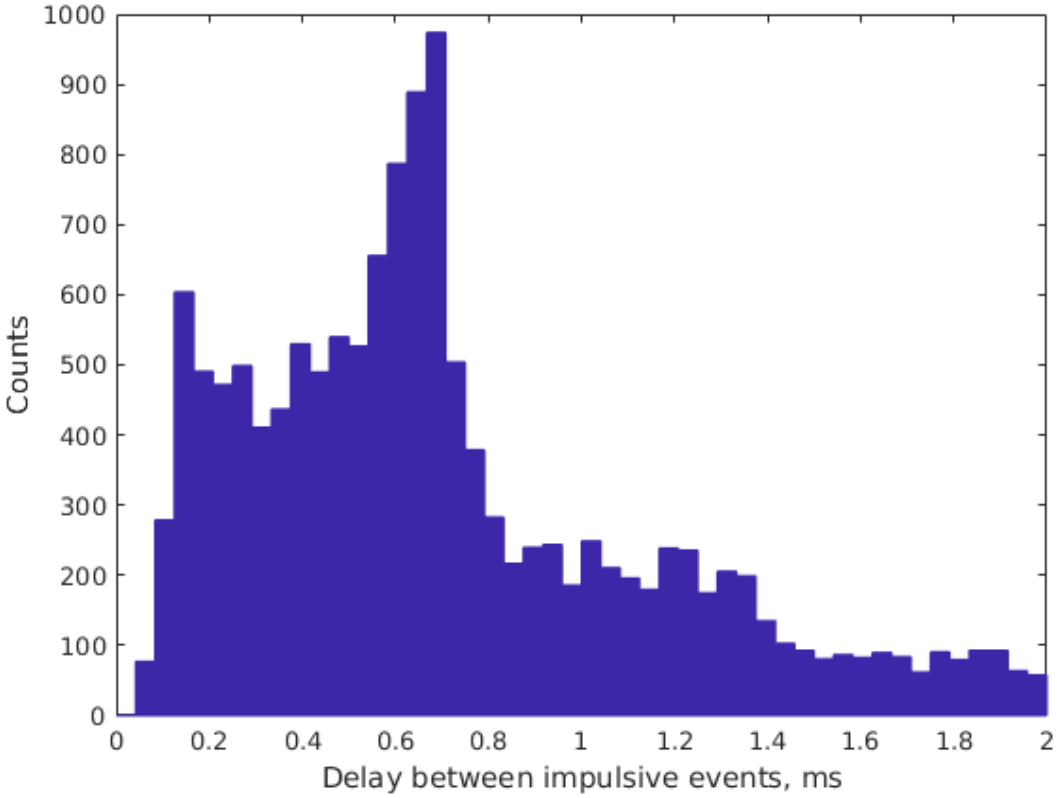


Figure 4.10: Histogram of the gap between pulse events. This source was weaker, resulting in more delay scatter than source A.

Ellingson, S. W., J. D. Bunton, and J. F. Bell (2000). "Cancellation of GLONASS signals from radio astronomy data". In: *Radio Telescopes*. Ed. by H. R. Butcher. Vol. 4015. Proceedings of the SPIE, pp. 400–407. DOI: 10.1117/12.390468. eprint: astro-ph/0002516.

Fridman, P. A. and W. A. Baan (2001). "RFI mitigation methods in radio astronomy". In: *Astronomy and Astrophysics* 378, pp. 327–344. DOI: 10.1051/0004-6361:20011166.

Glenn Jones, Wesley New (2013). *ROACH-2 Revision 2*. https://casper.berkeley.edu/wiki/ROACH-2_Revision_2.

Misra, S., R. D. De Roo, and C. S. Ruf (2012). "An Improved Radio Frequency Interference Model: Reevaluation of the Kurtosis Detection Algorithm Performance Under Central-Limit Conditions". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.11, pp. 4565–4574. ISSN: 0196-2892. DOI: 10.1109/TGRS.2012.2191972.

Mitchell, D. A. and J. G. Robertson (2005). "Reference antenna techniques for canceling radio frequency interference due to moving sources". In: *Radio Science* 40, RS5S11, RS5S11. DOI: 10.1029/2004RS003152. arXiv: 1003.5997 [astro-ph.IM].

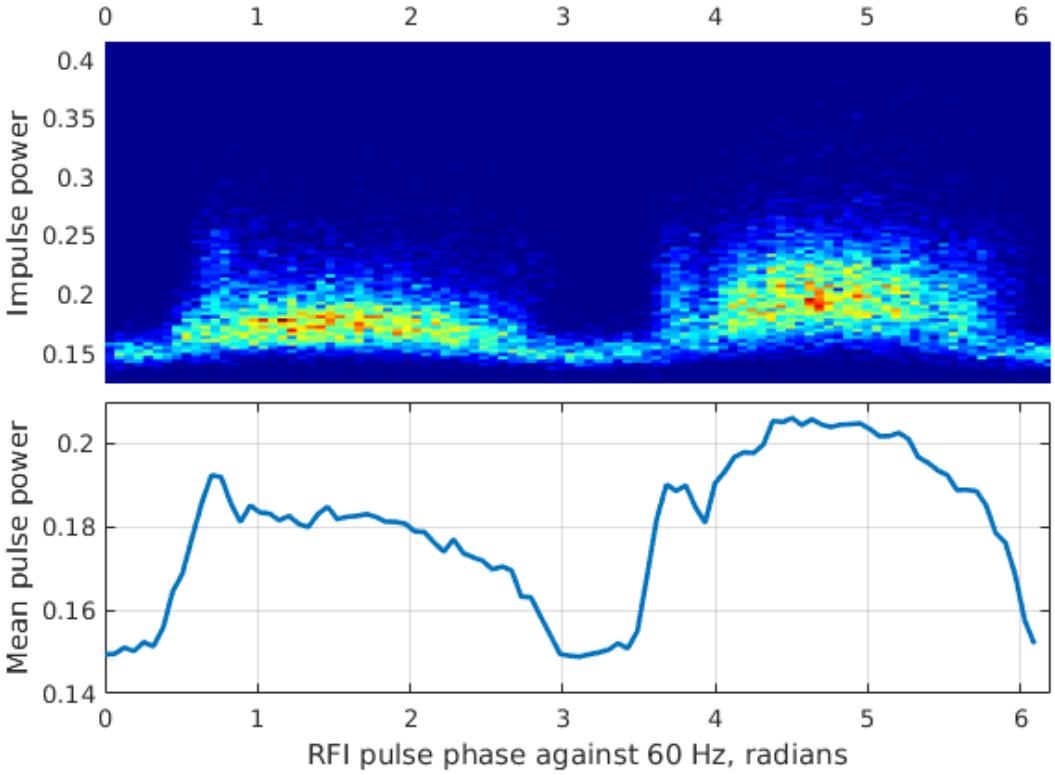


Figure 4.11: Top: Histogram of pulses against phase of a 60 Hz sinusoid for Source B. Bottom: Average power of detected pulses. Compare to Figure 4.7

Monroe, Ryan (2018). *Github Repository*. https://github.com/ryan-monroe/rfi_paper_2018.

Nita, GeluM. et al. (2007). "Radio Frequency Interference Excision Using Spectral-Domain Statistics". In: *Publications of the Astronomical Society of the Pacific* 119.857, pp. 805–827. ISSN: 00046280, 15383873. URL: http://www.jstor.org/stable/10.1086/520938.

Offringa, A. et al. (2015). "The Low-Frequency Environment of the Murchison Widefield Array: Radio-Frequency Interference Analysis and Mitigation". In: *Publications of the Astron. Soc. of Australia* 32, e008, e008. DOI: 10.1017/pasa.2015.7. arXiv: 1501.03946 [astro-ph.IM].

Offringa, A. R. et al. (2010). "A LOFAR RFI detection pipeline and its first results". In: *ArXiv e-prints*. arXiv: 1007.2089 [astro-ph.IM].

Raza, J., A. J. Boonstra, and A. J. van der Veen (2002). "Spatial filtering of RF interference in radio astronomy". In: *IEEE Signal Processing Letters* 9.2, pp. 64–67. ISSN: 1070-9908. DOI: 10.1109/97.991140.

Roo, R. D. De, S. Misra, and C. S. Ruf (2007). "Sensitivity of the Kurtosis Statistic as a Detector of Pulsed Sinusoidal RFI". In: *IEEE Transactions on Geoscience and*

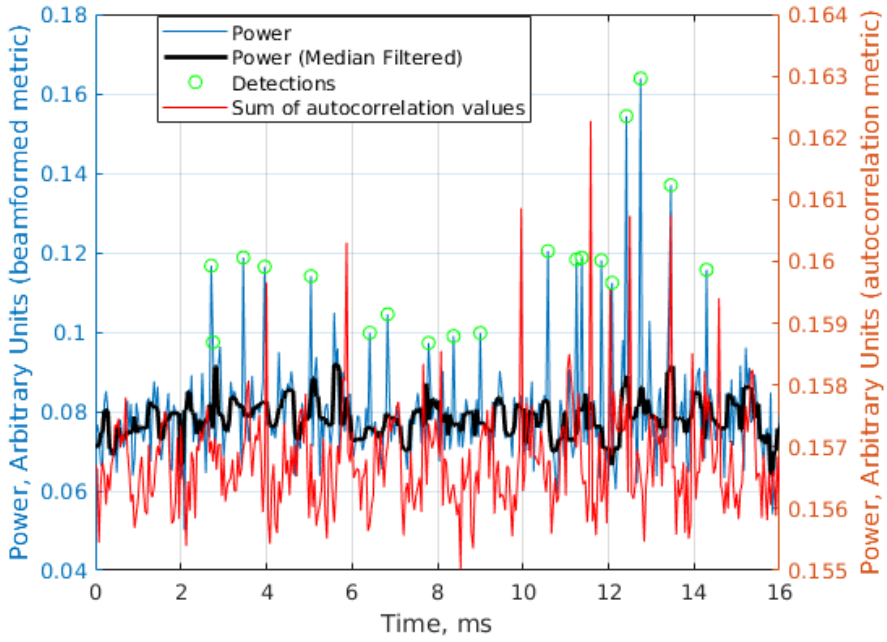


Figure 4.12: RFI detections, with an overlaid ~60 hz sine wave fit to event detections. (Source B).

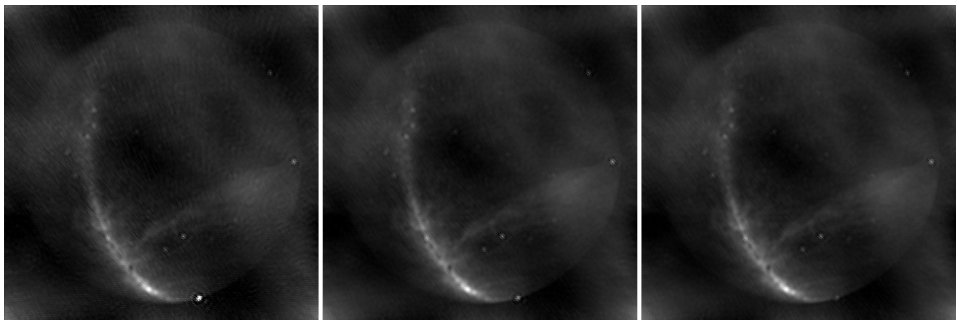


Figure 4.13: Results from RFI detector. Left frame:no flagging. Center frame: flagging based only on autocorrelation data. Right frame: beam-formed data for detection. The algorithm was seeking to mitigate the source near the bottom of the frame.

Remote Sensing 45.7, pp. 1938–1946. ISSN: 0196-2892. DOI: 10.1109/TGRS.2006.888101.

Chapter 5

SELF-TRIGGERED RADIO DETECTION AND IDENTIFICATION OF COSMIC AIR SHOWERS WITH THE OVRO-LWA

5.1 Introduction

The detection of radio signals has advanced to a standard technique in air shower arrays in the past decade. Dedicated arrays such as AERA (Aab et al., 2016a), as part of the Pierre Auger Observatory (Pierre Auger Collaboration, 2015) or Tunka-Rex (Bezyazeekov et al., 2015) regularly measure energy and composition. Radio detection of air showers combines the precision of optical methods with the almost continuous uptime of particle detectors (T. Huege, 2016a). Radio detection has shown to yield very precise composition measurements (Buitink, Corstanje, et al., 2014) and stable and excellent energy resolution (Aab et al., 2016b), which are the key parameters needed to understand the origin of cosmic rays.

Ground arrays rely on particle detectors to trigger on or identify cosmic rays from radio data of the antennas. Thereby, the efficiency of the radio detectors, which intrinsically increases for horizontal air showers as compared to particle detectors(Kambeitz, 2017), cannot reach its full potential, as the efficiency of the particle detectors is inherited through the trigger. Radio detection is expected to be more efficient because radio waves have a much longer range of propagation compared to the particles in the air shower, which are stopped after a few kilometers in the atmosphere. This is particularly a limitation for more inclined showers. Radio telescopes like LOFAR that are able to measure air shower signals (Schellart, A. Nelles, et al., 2013) have built particle arrays to facilitate the trigger (Thoudam et al., 2014). While the particle detectors also provide additional information for composition studies (Buitink and al., 2016), these additional hardware requirements complicate the search for cosmic ray signals in non-dedicated radio arrays and limit the efficiency of air shower detection.

Most experiments have chosen to rely on an external trigger due to the challenges inherent to a trigger on the radio pulse itself, a *self-trigger*. In most environments, simply triggering on pulses that are non-repeating and broadband by means of a threshold above the noise floor will yield extremely high trigger rates due to local

human activity. Only in remote regions of Antarctica, experiments aiming to detect neutrinos have successfully used an independent self-trigger to detect cosmic rays (Barwick et al., 2017; Hoover et al., 2010).

Early experiments to detect air showers have employed a self-trigger, but all ultimately had to verify their data using a particle array, partly due to limited knowledge of the nature of the radio emission pattern, and partly due to the complexity of finding a perfect discriminator (Ardouin et al., 2009; Schmidt, 2009; Abreu et al., 2012; Kelley, 2013). All working self-triggers had to severely limit the acceptance of signals due to local noise sources, such as transmitters, or airplanes (The Pierre Auger Collaboration, 2016). This can be overcome by using an external trigger, which was then deemed more reliable.

It has been proposed to built radio arrays to detect the radio signal following the interaction of a τ-neutrino of an energy above 10^{15} eV (Martineau-Huynh et al., 2017; Liu et al., 2017). In order to reach the necessary sensitivity to be able to detect these neutrinos that are expected in conjunction with the interaction of cosmic rays with the cosmic microwave background (Greisen, 1966; Zatsepin and Kuz'min, 1966), these experiments will need to employ a radio self-trigger (J. Abraham et al., 2008; Aartsen, K. Abraham, et al., 2016). If deployed as a distributed array without cabled connections between the antennas, the trigger has to be extremely effective at rejecting RFI, as data rates will be limited. Also, the computational power of autonomous stations will be restricted, which won't allow for complex procedures to identify the correct signals. It will, therefore, be necessary to know the requirements for a successful self-trigger when designing the detector system.

In addition to cosmic ray science, there has been an explosive growth in low-frequency arrays for 21 cm cosmology. These arrays, including LOFAR (van Haarlem et al., 2013), MWA (Tingay et al., 2013), PAPER (Parsons et al., 2010), and OVRO-LWA offer excellent collecting area and computing resources, but in many cases are limited by both their impulsive RFI[1] environment (ASTRON, 2018; Offringa et al., 2013), as well as the fact that their array configurations were not optimized for the task of cosmic ray detection. That said, accessing these arrays could provide an opportunity to perform a commensal cosmic ray science at low cost.

The OVRO-LWA provides the ideal testing ground for both the development of a self-trigger, and the task of re-purposing existing arrays towards detecting cosmic

[1] Radio Frequency Interference

rays. With few computational restrictions, different avenues can be explored and it can be shown whether air shower detection without a confirmation of particle arrays is at all feasible, let alone the radio detection of tau neutrinos outside of polar regions. Having all signals brought back to a central location (as opposed to beam-forming in the field, for instance), as well as possessing versatile digital back-end (Kocz, Greenhill, Barsdell, Price, et al., 2015), enable a wide variety of techniques to be quickly prototyped and validated. As well as being a powerful and commensal cosmic ray experiment, running as successful and efficient self-trigger at the OVRA-LWA can be a first step in proving the feasibility of proposed neutrino arrays and might be an important step towards an independent way of an air shower detection.

This paper will discuss the successful design of a cosmic ray self-trigger system using only the radio signal. First, we will summarize the technique used (Section 5.3), after which we will discuss the firmware and software algorithms used to perform RFI filtering (Sections 5.5, 5.6). A brief characterization of the types of RFI seen, as well as successful cosmic ray detections, are covered in Section 5.7. Recommendations for future instrument design and "Lessons Learned" are discussed in Sections 5.8 and 5.9, respectively. All software and firmware used in this work are available at (Monroe, 2018).

5.2 The Owens Valley Radio Observatory Long Wavelength Array

This experiment was performed by re-purposing an existing instrument: the Owens Valley Radio Observatory Long Wavelength Array (OVRO-LWA) to perform cosmic ray detections without modifying any hardware. The system was originally designed as a fully cross-correlating interferometer, for which it continues to be used in its primary purpose of imaging, transient science, and cosmology. As such, the array is at a site which has no particle detectors or fluorescence telescopes (the standard methods of detecting cosmic ray events).

The OVRO-LWA is located at the Owens Valley Radio Observatory, near Bishop, CA, where it has substantial shielding from major population centers, courtesy of the Sierra and Inyo mountains. As a consequence, the environment is very good for intentionally transmitted RFI. However, power lines in the valley are responsible for impulsive, band-limited RFI which makes the detection of cosmic rays challenging. Emphasizing cosmology and transient science, the array consists of 288 dual-polarization antennas (of which only 256 are in active use) distributed

across a 1.4 km region. All but 37 of these antennas are randomly distributed across a 200 m diameter core, which was the focus of this effort. OVRO-LWA inherits design for many components from the New Mexico Long Wavelength Array, which is similar (but with solely a smaller, 100 m diameter core, and narrow-band software correlation/beamforming instead of full cross-correlation). Each antenna sees nearly the entire sky, with a 3 dB point of approximately 40° elevation angle. At OVRO-LWA, each antenna is baseband sampled at $F_s = 196.608$ MSPS, for a Nyquist frequency of ~ 98.3 MHz. This is performed with 8-bit ADC16x250-8 Analog to Digital Converter (ADC) boards (Dexter and al., 2014), each an on-board HMCAD1511 (Devices, 2015) ADC. In the band of interest, both the power observed by each antenna and the system temperature are dominated by the galactic background. Signals are then processed by 16 CASPER ROACH2 FPGAs, each of which contains a Virtex 6 XC6VSX475T FPGA (32 inputs or 16 antennas per FPGA). During normal correlator operations, these FPGAs perform the "F-engine" (Eastwood et al., 2017; Anderson et al., 2017; Kocz, Greenhill, Barsdell, Bernardi, et al., 2014) portion of the correlator. Subsequently, a 10 GbE / 40 GbE Ethernet switch is used to transfer data to a set of GPU computers for further processing, which is not relevant for this work.

5.3 Summary of Technique

For this experiment, we re-use the existing hardware in its current configuration for the task of cosmic ray detection. The standard F-engine firmware is replaced with firmware which is designed to detect impulsive events while identifying and filtering known RFI sources. Preliminary detections are made at an individual FPGA level: initially, each FPGA analyzes data from the 16 antennas it services (which are spatially localized to rectangular regions roughly 25 m by 100 m in extent), and makes a preliminary decision using only this data. After detecting a promising impulsive event, the FPGA transmits an appropriate time-stamp to all 15 other FPGAs, which respond by dumping their data for that time window over 10 GbE to a host PC. That server stores the candidate to disk for offline processing. After on-chip RFI filtering, the trigger is sensitive enough that 60% (typical conditions) to 80% (ideal conditions) of on-chip triggered events are caused by random thermal noise – evidence that the system is close to thermally limited.

Candidate events are then processed in MATLAB to produce the higher level statistics necessary for identification of true cosmic ray events. Primary flagging metrics include observed power at each antenna, direction, and range of source arrival,

quality of a spherical wavefront model in source fitting, and time-domain clustering of events. This flagging reduces the number of impulsive candidates by a factor of ~220,000. In the 40-hour dataset which was collected, 16 events are sufficiently promising to warrant human inspection. Of these, 9 events (56%) appear to be caused by cosmic rays, a further 3 events (19%) are unable to be confirmed or rejected as astrophysical events and 4 events (25%) are most likely caused by RFI. This is equivalent to 5-6 cosmic-ray driven events per day under good RFI conditions.

The largest challenge in this project is filtering RFI events from cosmic ray events: under good RFI conditions, the system captures approximately 400K impulsive candidates for every cosmic ray (~100 Hz, versus ~hourly). This issue is worsened by the fact that this system was not designed with cosmic ray detections in mind. There is insufficient communication bandwidth between the FPGAs and the Ethernet switch with which they are connected, making transmission of the entire raw ADC data-stream impossible. For this reason and others, triggering must be performed on an individual FPGA-basis with minimal communication, severely limiting the systems ability to reject RFI.

5.4 Identification of cosmic rays

Although the RFI conditions of the OVRO-LWA site are very good by interferometric standards, the impulsive RFI environment is very challenging. A typical spectrum as seen by the OVRO-LWA can be found in Figure 5.1. Approximately 500 impulsive RFI events (with more power than the galactic background noise) occur per second, compared to the cosmic ray incidence rate of ~hourly (Corstanje et al., 2011; Aartsen, Abbasi, et al., 2013).For this reason, aggressive use of descriptive statistics must be used to filter out promising cosmic ray candidates from spurious events. Broadly speaking, these properties can be described as:

- Input signal is roughly a band-limited impulse (A. Nelles, Buitink, Corstanje, E. Enriquez, et al., 2013), convolved with the instrumental impulse response
- Time and direction of arrival is usually not coincident with a known or detected RFI source (mostly stationary sources and airplanes), which can be clustered in either direction or time+direction
- Power distribution in the axis of ray travel is similar to that described in (A. Nelles, Buitink, Corstanje, J. E. Enriquez, et al., 2015) – for this work, we

searched for a localized region of power with a FWHM[2] of 100-200 m on the ground, even though many cosmic rays could have a broader extent

- Polarization properties roughly match what is expected given an events direction of arrival and the local geomagnetic field (Schellart, Buitink, et al., 2014)
- Finally, many events are discarded if it is likely that event power is insufficient for quality parameter estimation, or instrumental errors (such as dropped data due to network bandwidth saturation) are likely to result in ill-conditioned fits

The system described below was designed to use these properties with the goal of both detecting cosmic rays while rejecting as many RFI events as possible, as early as feasible in the signal chain so as to manage data-rates and make more sensitive detections.

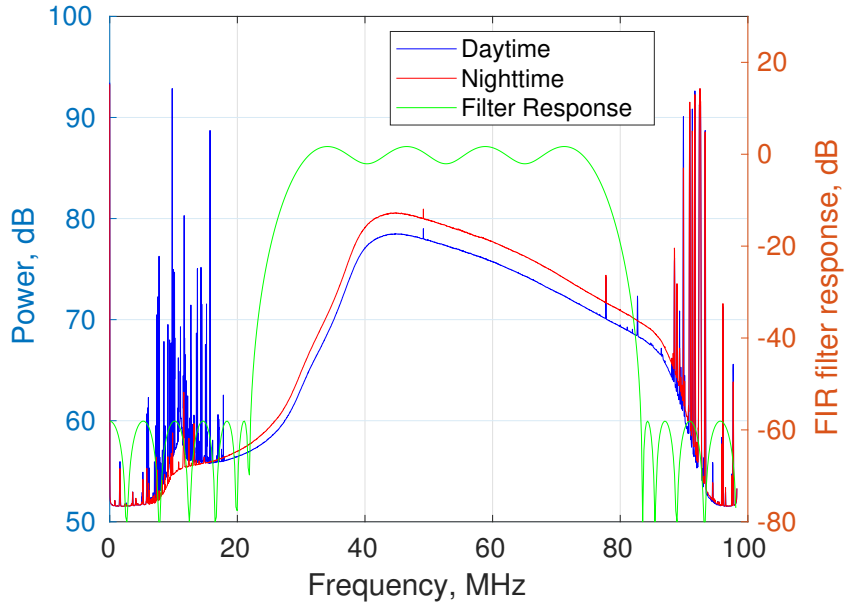


Figure 5.1: Typical daytime spectrum as seen from the OVRO-LWA. The on-chip integrators used to produce this spectrum saturate at 93 dB, meaning that the true power of the RFI at high frequencies may be higher than what is shown in this figure. Also shown is the FIR filter response used in Sections 5.5 and 5.6.

[2]Full-Width-Half-Max

5.5 Trigger

Here we discuss the algorithms used in the FPGA firmware to detect cosmic rays and perform preliminary RFI filtering. For definitions of symbols used in this section, refer to Table 5.1.

Symbol	Meaning
k	Input index ($0 \sim 31$ or $0 \sim 511$)
$v_k\,[n]$	Sampled voltage for k'th input
$v_{filt,k}\,[n]$	Filtered voltage for k'th input
$h\,[m]$	Coefficients for 40-tap symmetric FIR filter
$p_k\,[n]$	Filtered power for k'th input
$p_{sm,k}\,[n]$	Filtered power for k'th input (smoothed)
T_k	Triggering threshold for k'th input
$E_{I,k}\,[n]$	Input-wise event for k'th input
D_k	Corrective delay for k'th input (in samples)
$E_{D,k}\,[n]$	Event for k'th input (delayed)
N_{bl}	Number of samples to block for consecutive pulses
$E_{B,k}\,[n]$	Event for k'th input (after repeated event blocking)
N_{sus}	Event sustaining window (in samples)
$E_{sus,k}\,[n]$	Event for k'th input (additionally sustained)
$N_{trig}\,[n]$	Number of input triggers required for FPGA trigger
$C\,[n]$	FPGA-wise event detection

Table 5.1: Firmware symbol definitions.

As mentioned in Section 5.3, each ROACH2 processes 16 antennas, for 32 inputs per FPGA. A circular buffer stores raw ADC samples for triggered dumping, while the signals are processed in parallel to detect impulsive events received by the system. A simplified block diagram of the firmware is shown in Figure 5.2. All math used in this work is real-valued.

Filtering and power smoothing

First, a 40-tap symmetric-coefficient bandpass FIR filter (5.1) is used to filter the ADC voltage samples to the range [30,75] MHz (Figure 5.1), which was chosen to avoid narrow-band, high-duty-cycle RFI sources. Xilinx FPGAs have special features which half the filter resource requirements for symmetric FIR filters, driving that decision. Afterward, filtered voltages are squared to compute power (5.2), and subsequently smoothed with a 4-tap running average filter (5.3); the value of 4 samples (20 ns) over the band-limited impulse time of 7 samples (35 ns) in this system was chosen to prioritize accurate time-of-arrival estimation over maximum sensitivity, justified because downstream processing could not handle such low SNR

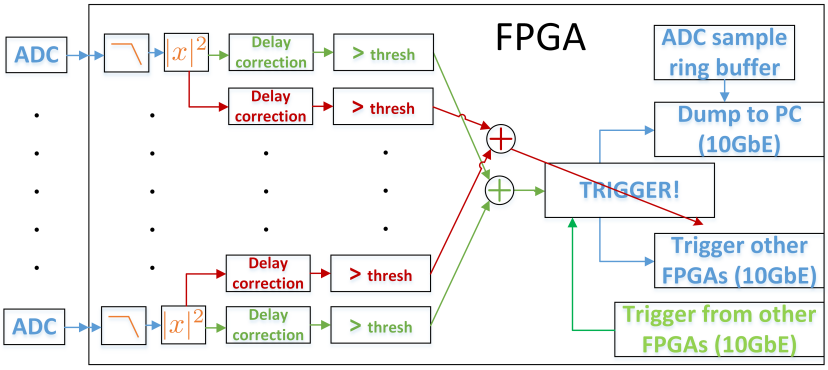


Figure 5.2: Simplified block diagram of the FPGA system.

events anyways. This smoothing and triggering process does not explicitly filter for impulsive events, but the short integration length prefers over events with similar power levels but longer duration (and therefore, lower instantaneous power).

Thresholding and RFI mitigation

Each of these signals which exceeds a threshold (5.4) is registered as an ADC-wise event. Cable length-delays are then removed from each signal using pre-computed values (5.6). In order to facilitate a primitive time-of-arrival estimate on-chip, all ADC-wise events which occur within N_{bl} of a preceding event are suppressed: otherwise, a powerful event could trigger the RFI blocking system (Section 5.5) multiple times, resulting in powerful events being blocked regardless of incident direction (5.7); N_{bl} is only non-zero in the directional RFI blocker, as discussed in Section 5.5. Since events will not typically be detected by all ADCs simultaneously even after cable delay[3], event detections are sustained for an appropriate number of samples (5.8) N_{sus}. This value was set to 140, but could have been set as small as $100 * {}^{F_s}/_{c} \approxeq 66$ samples, where $c = 3e8\,{}^{\mathrm{m}}/_{\mathrm{s}}$ is the speed of light, and 100 m is the maximum extent of the core OVRO-LWA antennas that a single FPGA services. If after thresholding, suppression and sustaining, more than N_{trig} ADCs have sustained events at any given time, that FPGA is said to have made an impulsive event detection (5.9).

Trigger and dump via Ethernet

The time-stamp – which is synchronized by a global 1PPS[4] signal – is then transmitted to all other FPGAs via a 10 GbE interface, which synchronously halt writing

[3] Due to geometric effects
[4] 1 Pulse Per Second

to their circular buffers and transmit an event record (meta-data as well as 4096 raw ADC sample – equivalent to 20 μs near the trigger time) to the host PC, which stored the events using the "tcpdump" tool. A logical "Rising Edge" operation is applied to the signals after thresholding individual ADC detections to facilitate the RFI blocking algorithm (which will be discussed below, in Section 5.5). There was 2.6 ms of dead time after each trigger (arbitrarily chosen, this could have been made shorter).

FPGA signal processing: summary

The operations described above can be approximately summarized with the following equations:

$$\text{(bandpass filter) } v_{filt,k}[n] = v_k[n] * h[m] \quad (5.1)$$

$$\text{(compute power) } p_k[n] = v_{filt,k}[n]^2 \quad (5.2)$$

$$\text{(smooth power) } p_{smooth,k}[n] = \sum_{m=0}^{3} p_k[n-m] \quad (5.3)$$

$$\text{(threshold) } E_{I,k}[n] = p_{sm,k}[n] > T_k \quad (5.4)$$

$$\text{(rising edge) } E'_{I,k}[n] = E_{I,k}[n] \,\&\neg E_{I,k}[n-1] \quad (5.5)$$

$$\text{(delay) } E_{D,k}[n] = E_{I,k}[n-D_k] \quad (5.6)$$

$$\text{(block repeat events) } E_{B,k}[n] = E_{D,k}[n] \,\&\neg \max_{s\in\{1\ldots N_{bl}\}} \left\{E_{D,k}[n-s]\right\} \quad (5.7)$$

$$\text{(sustain detections) } E_{sus,k}[n] = \max_{s\in\{0\ldots N_{sus}+1\}} \left(E_{B,k}[n-s]\right) \quad (5.8)$$

$$\text{(count events) } C[n] = \left(\sum_{k=0}^{31} E_{I-dl,k}[n]\right) > N_{trig} \quad (5.9)$$

Practical considerations

There are some run-time practicalities to making this system work. The thresholds T_k must be scaled to account for differences in system gain across antennas and polarizations. This is done once at startup by selecting a random selection of 4096 samples samples from each of the 512 array inputs, and enforcing that each T_k is inversely proportional to the power in that input: $T_k = G\left(\sum_{n=0}^{4095} v_k[n]^2\right)^{-1}$. As background RFI event rate varies, G is adjusted slowly (specifically, scaled by a factor of 2%∼50% every six seconds, depending on how far away from the target range the current dump rate falls) such that the typical number of detected events

falls within a user-defined region (12-50 events/sec in this case: 80% of those are typically random thermal noise, whereas almost all of the remainder are RFI). The upper bound was set by the computational power available for post-processing, while the lower bound was set as a quarter of the upper bound to avoid rapid adjustments in threshold values.

On-chip RFI suppression

As the system design was iterated, it was found that the large number of RFI events was outstripping the processing ability of the software system. In order to mitigate this problem, a parallel (on-FPGA) algorithm was designed to block incident RFI coming from particular directions – this strategy was feasible because almost all of the RFI came from only three distinct, stationary sources (see Section 5.7). This was implemented by replicating equations (5.6) through (5.9) three times, gating the main detection on a lack of activity on these "RFI blockers". Each of these directional blocking modules is configured with delays describing the timing of signals with regard to their arrival at the FPGA (equivalent to the sum of aforementioned cable length delay and the geometric delay appropriate for sources in the given direction). A detection with one of these modules blocks an FPGA trigger for the past and future $\sim 2.5\,\mu s$. A few minor departures from the original event trigger are also necessary for the RFI blocker: first, the original sustaining period of $N_{sus} = 140$ samples was designed to capture any impulsive events: for the RFI blocking module, this is adjusted to $N_{sus} = 6$ samples, a number which (per captured data) allows the RFI system to detect the majority of events from the chosen source, while blocking events from only a trivial fraction of the sky. Additionally, (5.5) and (5.7) were added, which temporarily suppress subsequent detections – transforming the detectors from simple power thresholds into a measure of time-of-arrival (otherwise, powerful events would trigger the RFI detectors multiple times, causing these powerful events to be blocked regardless of direction).

This blocking technique is imperfect: random thermal noise would occasionally trigger the RFI detector of individual inputs shortly before the arrival of a true RFI event. Per (5.7), this would preclude that input triggering at the appropriate time for the true event, allowing sufficiently weak RFI events (such that it was seen by a marginal number of RFI detection inputs) to slip this filter. The RFI blocker suppresses roughly 315/316 events which would otherwise trigger the system, allowing the detection threshold to be reduced by a factor of roughly 2.8, for the same candidate event rate (Figure 5.3). Although many RFI events are not blocked by this

filtering mechanism, this filtering mechanism the remaining events can be handled in subsequent software post-processing. A more effective RFI filtering technique is described and prototyped in Section (5.9).

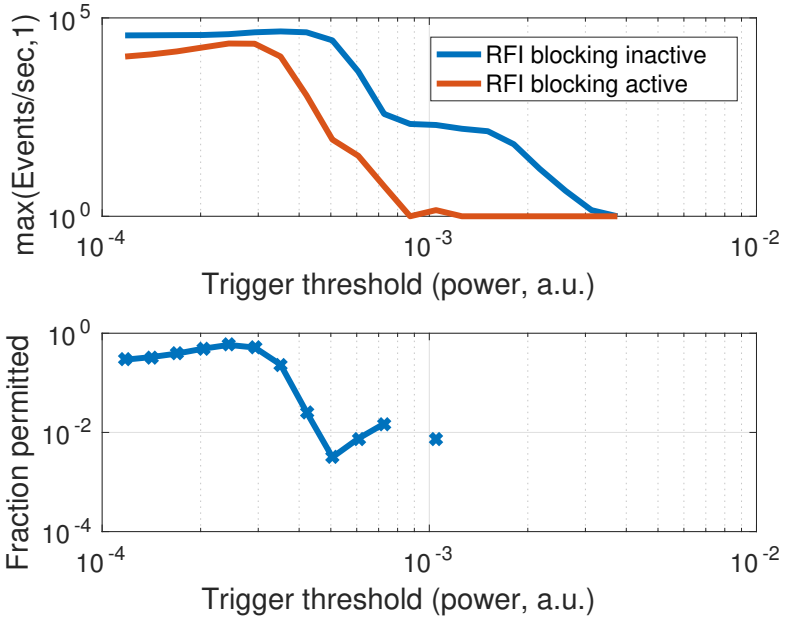


Figure 5.3: Top: event rates both with and without on-chip filtering applied. Each data point was collected at a distinct time in a time-variable RFI environment, which limits interpretability of the data. Bottom: approximate fraction of events which remained after filtering.

5.6 Analysis

Here we discuss the software algorithms used to detect cosmic rays. For definitions of symbols in this section, refer to Table 5.2.

Filtering and time-of-arrival estimation

After receiving event data from the FPGAs, candidate events are then processed to produce higher-level statistics. A simplified summary is shown in Figure 5.4. For each record, ADC samples are re-filtered as in the FPGA, reproducing $p_{smooth,k}[n]$. The temporal center of the event is estimated as in (5.10): all subsequent analysis is constrained to 250 samples (1.2 μs) centered around t_{center} to minimize the effects of noise and non-impulsive events in detections. Additionally, an estimate of the background noise in each input is made by computing (5.11): the mean of several

Symbol	Meaning
$F_s = 196.608\,\text{MHz}$	ADC sampling rate
t_{center}	Estimate of event arrival time, in samples
$k \in \{0..511\}$	Input index (note: across entire array)
$p_{sm,k}\,[n]$	Filtered power for k'th input (smoothed)
N_k	Noise estimate for k'th input
P_k	Event power estimate for k'th input
t_k	time of arrival estimate for k'th input
d_k	Distance from the k'th inputs antenna to event
flag_k	Flags discarding the k'th input
$S = \left[S_x, S_y, S_z\right]^\top$	Position of event source
$A_k = \left[A_{x,k}, A_{y,k}, A_{z,k}\right]^\top$	Position of antenna serving k'th input
G_k	Estimate of relative sample delay implied by fit
$R\,[n]$	Estimated pulse power profile

Table 5.2: Software symbol definitions.

samples of $p_{smooth,k}$ far[5] from t_{cent}. The event power on each input is estimated with (5.12). Any of the 512 inputs with an estimated Signal to Noise Ratio (SNR, estimated as the ratio between input event power and background noise power) less than 5 is flagged (5.14) and ignored for further processing – any event with fewer than 50 un-flagged inputs is rejected as a systematic or otherwise too weak for analysis. Event time of arrival for a given input is estimated as (5.13). Flagging is done at this stage because the arg-max function will produce a "time-of-arrival" regardless of the quality of the input signal: subsequent stages would otherwise fail due to a large number of outliers contaminating the true signal.

Direction of arrival fitting

The next step of processing fits for source location. Knowing the positions of each antenna as well as the time of arrival at those antennas (as well as making a spherical wavefront assumption) allows us to treat each input as a measurement of the relative distance from the antenna to the source. Estimates of cable length are subtracted from the data and the resultant times are converted into units of meters traveled at the speed of light by multiplying by c/F_s. As a crude outlier rejection, a robust plane fit is performed on the tuples (X_k, Y_k, d_k)[6], using the MATLAB function "pcfitplane" (Inc., 2015b; Torr and Zisserman, 2000). Inputs which fell sufficiently far from the robust plane fit were also flagged – the threshold was set such that events which were

[5]Because it was before the triggering time, events only rarely fall near these samples

[6]The altitude of the antennas is ignored for this step: the array is very nearly planar

farther than the edge of the array were not impacted. The 3D position of the source is then estimated as (5.17), using the planar fit as a starting point. The search was performed in the modified spherical coordinates $S_{az}, S_{el}, \log_{10}(S_d)$, where source positions were converted to Cartesian as needed – this had empirically better results with the optimizer. Sequential 1D global searches were done in azimuth, elevation and $\log_{10}$ (range) (three passes), followed by a call to the MATLAB tool "fminunc", which uses a quasi-newton solver (Inc., 2015a). Additionally, by aligning all un-flagged power signals using the geometric delays implied by the source position (5.18) and summing across non-flagged inputs, a pulse profile for the event is constructed (5.19) for visual inspection. The fruits of this effort are the key metrics by which candidate cosmic ray events are automatically filtered: azimuth, elevation, range, fitting residual, $N_{inputs-detected}$, peak pulse power.

Software processing: summary

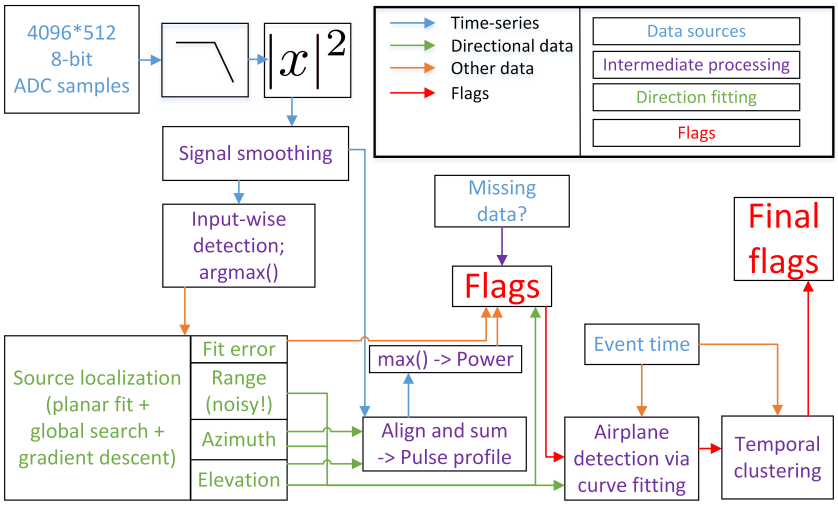


Figure 5.4: Simplified block diagram of the software processing pipeline.

Most of the process described above can be summarized as

$$\text{(global time estimate) } t_{center} = \arg\max_n \sum_{k=0}^{511} p_{sm,k}\,[n]^2 \quad (5.10)$$

$$\text{(input noise estimate) } N_k = \frac{1}{301}\sum_{n=100}^{400} p_{sm,k}\,[n] \quad (5.11)$$

$$\text{(input power estimate) } P_k = \max_n \left(p_{sm,k}\,[n]\right) \quad (5.12)$$

$$\text{(input TOA estimate) } t_k = \arg\max_n \left(p_{sm,k}\,[n]\right) \quad (5.13)$$

$$\text{(input flag by SNR) flag}_k = {}^{P_k}/_{N_k} < 5 \quad (5.14)$$

$$\text{(fit plane) } [\sim, \text{inliers}] = \text{pcfitplane}\,(t_k, A_k \forall k | \neg\text{flag}) \quad (5.15)$$

$$\text{(flag on fitting error) flag}_k = \text{flag}_k | \neg\text{inliers} \quad (5.16)$$

$$\text{(fit location) } \{S, r\} = \arg\min_{S,r} \sum_{k|\neg\text{flag}} \left((S-A) - {}^{t_k \cdot c}/_{F_s} - r\right)^2 \quad (5.17)$$

$$\text{(estimate delays) } G_k = (S-A)\frac{F_s}{c} \quad (5.18)$$

$$\text{(estimate pulse profile) } R\,[n] = \frac{\sum_{k|\neg\text{flag}} p_k\,[n - G_k]}{\sum_k \neg\text{flag}} \quad (5.19)$$

RFI detection and flagging

Remembering that there are hundreds of thousands of RFI events for every true cosmic ray, the descriptive statistics from Section 5.6 must be used aggressively to remove non-astronomical events. The median high-SNR event has an RMS time-of-arrival residual across inputs of ~ 3 samples (15 ns, about half of which is due to signal processing losses). The distribution is bimodal, however, with a number of events having SNR sufficient for detection, but not localization with this algorithm (Figure 5.5). For this reason, events with $\sim$ 12 times the median fitting residual are discarded. Additionally, events which were only seen on a small number of inputs (for ethernet bandwidth or signal-strength reasons) are flagged in fear of an ill-conditioned direction fit. Of those that remain, most are clustered in a few azimuthal directions: these are likely RFI. Previous work (Chapter 4) has found these to likely be coronal arcing events from high-voltage power lines. Additionally, most RFI events are found to come from a low elevation angle, and some events can be localized to a short distance from the array. For this reason, events coming from a few tightly-constrained directions in azimuth (97° out of a total of 360° in azimuth, or a $\sim$ 27% of all directions), or low elevations ($< 2°$) and ranges (<1 km), are filtered, removing all but one event per 10,000.

Airplanes: an irritating source of RFI

An overwhelming majority of the events which remain after the aforementioned filtering steps draw smooth curves in the {azimuth, time} and {elevation, time} spaces. These are suspected to be airplanes, which can also be seen in OVRO-LWA imaging data. These events are additionally densely clustered in time: almost all remaining RFI events can be removed by flagging time epochs which contain more than 3 un-flagged events within a 100 second window (or in an adjacent 100 second window). At nighttime, this results in minimal (<5%) of time epochs being flagged, but during the daytime, heavy air traffic results in up to 37% of the data being flagged due to overhead air traffic. In general, these airplane-driven events do not interfere with the FPGA trigger beyond increasing the output data rate slightly.

After successful detection of cosmic rays using the cluster flagging technique, it was concluded that simply flagging in time was wasteful in terms of observing time. A more precise flagging of airplanes was achieved by performing local and robust curve fits to the {azimuth, time} and {elevation, time} spaces. Starting with the epoch containing the largest number of un-flagged events, these curves are fit to the data, predicting the path that the airplanes are taking over the array. Any events appearing in a direction close enough to the predicted path were flagged. Additional curves were added until no temporal epoch contained more than a set event rate. This resulted in approximately the same number of false positives but much more retained daytime data. Figure 5.6 shows a scatterplot of event arrival times, featuring many overhead airplanes, whereas Figure 5.7 shows the results of the curve-fitting process. Although providing a single statistic which summarizes this technique is challenging, the results can be visualized with Figure 5.8. One especially clean dataset can be visualized in Figure 5.9, and the full process of the entire automated RFI mitigation process can be visualized in Figure 5.10.

Manual inspection of promising events

Many un-flagged events were obviously associated with airplanes by visual inspection of {azimuth, elevation, time} scatter-plots, but were missed by flagging algorithms due to a variety of reasons – these were flagged by hand. Those few events which remain are visually inspected for a Gaussian-shaped spatial power distribution on the array, as well as a band-limited pulse profile and a polarization signature appropriate for the direction of arrival and Earths' local magnetic field. About two thirds of the events which reach this stage appear to be cosmic rays, and all of those with an appropriate power distribution as discussed in Section 5.4 obey

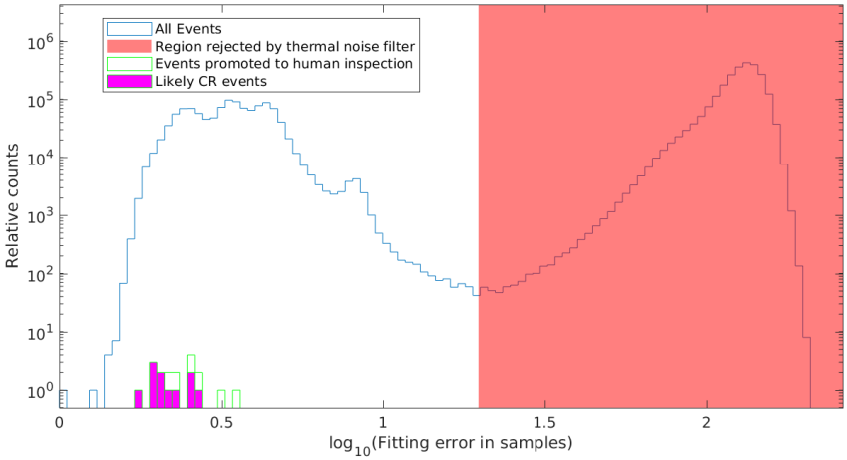


Figure 5.5: Histogram of residuals from the direction of arrival fitting. High fitting errors usually indicate a spurious trigger caused by random thermal noise.

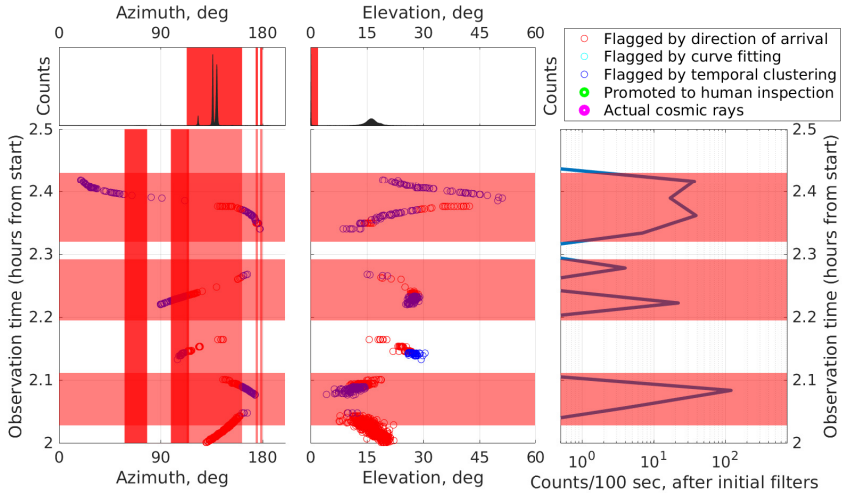


Figure 5.6: Upper left, upper center: histogram of events as a function of Azimuth (degrees North of East) and Elevation (degrees), respectively. Lower left, lower center: scatter-plots of the same against time. The many smooth traces are likely RFI originating from (or reflected by) planes. This dataset is from midday when airplanes were especially prominent: they were mostly filtered using the time-binning technique mentioned in Section 5.6. The vertical red bars are sections of azimuth and elevation which are automatically flagged. Red circles indicate events which pass all automated filters and must be investigated by humans. Despite the filtering, almost all remaining events come from airplanes. Time-domain clustering of these overhead events results in the flagging of a substantial portion of the dataset. Being taken during daytime, this is one of the worst datasets (from an RFI standpoint) collected.

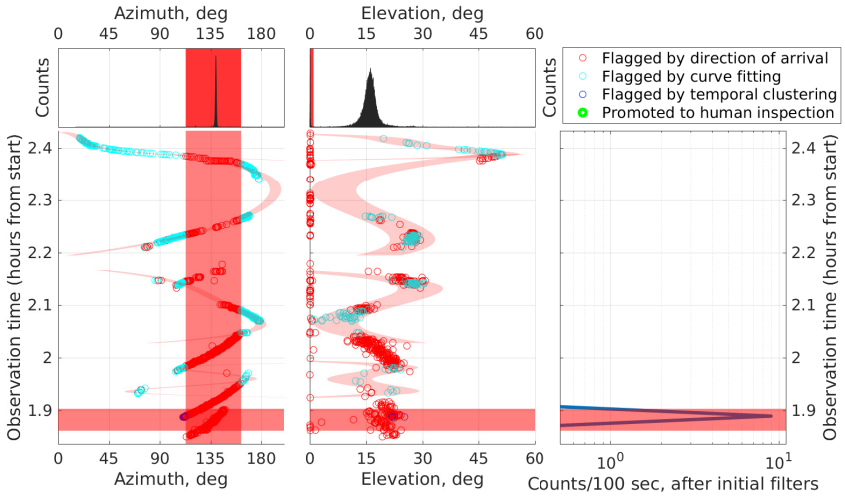


Figure 5.7: The same data as in Figure 5.6, with the additional application of a curve-fitting algorithm for airplane-driven RFI flagging. The quantity of data flagged is substantially reduced with little change in the false positive count of this stage of processing.

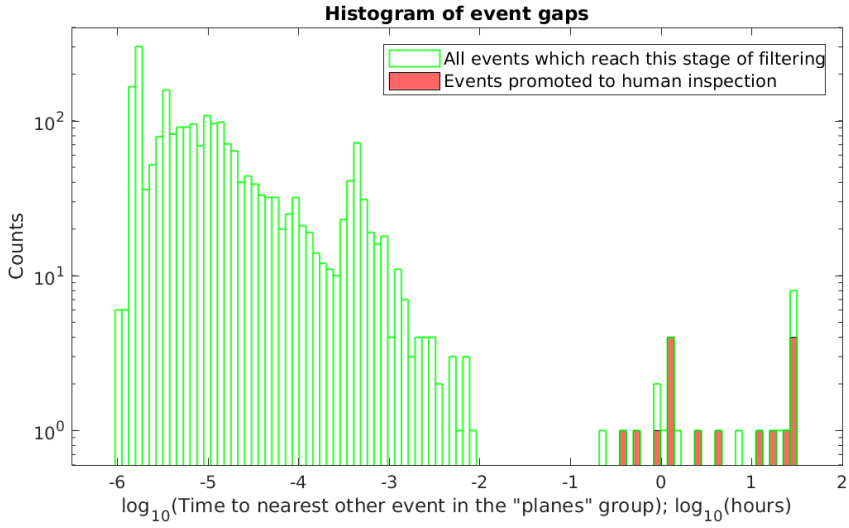


Figure 5.8: Another perspective on the airplane flagging routine, which is performed between the Upper Right and Lower Left plots of Figure 5.10. The airplane fitting algorithm cannot be reduced to a single monotonic decision statistic, so this histogram imperfectly visualizes those cuts.

every other detection metric. The level of similarity of these events which comply with our expectations for cosmic rays, combined with the lack of unflagged events which compellingly demonstrate some traits, while lacking others (excluding very weak events, of course), is strong evidence that these are cosmic ray detections and not false positives. All of the filtering thus far has been unbiased: Broadly speaking,

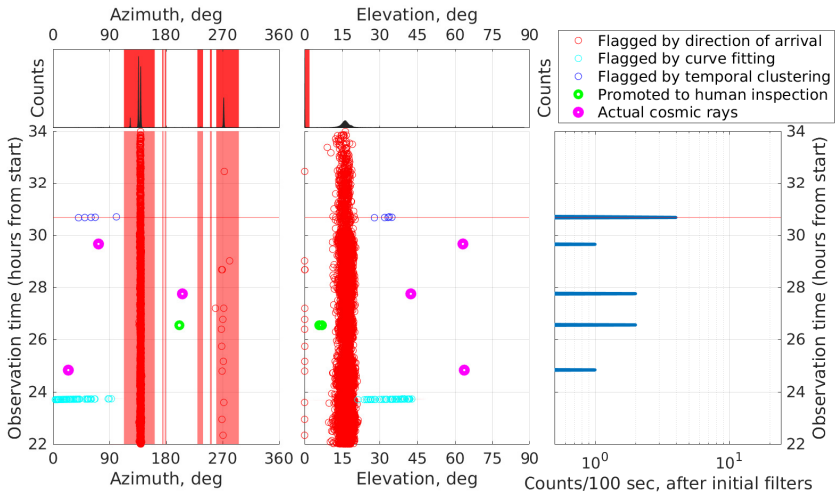


Figure 5.9: This dataset was one of the cleaner recorded. All events in magenta are likely to have originated from cosmic rays.

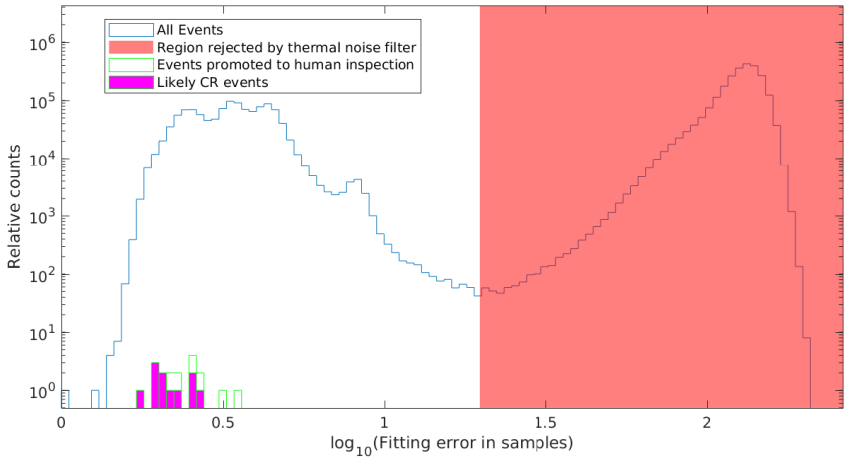


Figure 5.10: Another perspective of the event identification process. Each point is an event, placed at the appropriate location for its azimuth and zenith angle (North is up). Each panel of this figure represents a different stage of flagging. Likely cosmic ray events are given a black backdrop. **Upper-left:** All 473K events in the dataset which pass instrumental and event power checks. Events marked red are flagged based on azimuth and elevation cuts. **Upper-right:** Point color indicates observation time of arrival. Curves of matching color correspond to likely correlated airplane events. **Lower-left:** Event colors same as before– remaining temporally clustered events are flagged regardless of direction of arrival (they still appear to have time-direction structure, but were able to defeat the airplane-fitting routine). **Lower-right:** Events which are promoted to human inspection

we have uniformly discarded events based on arrival time and direction. Doing so reduces the dataset from ~500k events to 16 candidates. Going farther requires

inspecting the actual event data, which runs the dissatisfying risk of biasing results. Is a local bump in received power across the array a cosmic ray detection or just random thermal noise? For example, the methodology presented in this work will be likely to be more sensitive to events arriving from a high elevation angle, since the power distribution across the array of these high elevation angle events can be easily seen despite the small spatial extent of the array. OVRO-LWA does have an extended 1.4 km component to the array, but integrating it into the analysis is beyond the scope of this work. All events which are promoted to human inspection can be summarized in Table 5.3, or viewed in detail in the Appendix.

An attempt at automating this manual inspection process is presented in Figure 5.11. The main emphasis of manual inspection of candidate events was to confirm the presence of a spatial power distribution which was appropriately sized and oriented for a cosmic ray event. In an attempt to automate this, a constrained Gaussian + uniform background power model was fit to the spatial power profile of the event on the array for each event. The fraction of total power which was captured by the Gaussian is taken to be a measure of how much this event matches the distribution for a cosmic ray. This was coupled with a measure of the polarization agreement with theory. Each OVRO-LWA antenna contains one dipole each oriented in the N-S and E-W direction. Using the Gaussian fits described above, the ratio between the power captured in the N-S dipole as a ratio of the summed power captured in both Gaussians, is used as a measure of polarization. This is compared to the expected fractional power received in the N-S dipole, modeled by assuming that all power comes from the geomagnetic $v \times B$ emission. A primitive beam model of the OVRO-LWA antennas is used to predict the response in each dipole due to geomagnetic emission from the direction in question. Comparing prediction to reality allows for a measure of the similarity of the polarized signal to expectation. This statistic is limited because it does not model Askaryan emission, and is additionally subject to errors in power estimation, direction of arrival and antenna beam properties. In Figure 5.12, we show the actual polarization measure for sources coming from three distinct azimuth angles which show exceptionally large amounts of RFI. Each of these plots is expected to mostly contain events coming from a unique physical source: variation in the distribution of polarization measures can be used as an indicator of the quality of this measure in practice, which is to say accurate to about 2-3 percent. The remainder is likely due to imperfect modeling or systematics, especially for events originating from directions near parallel to the local geomagnetic field lines, or events with low received power.

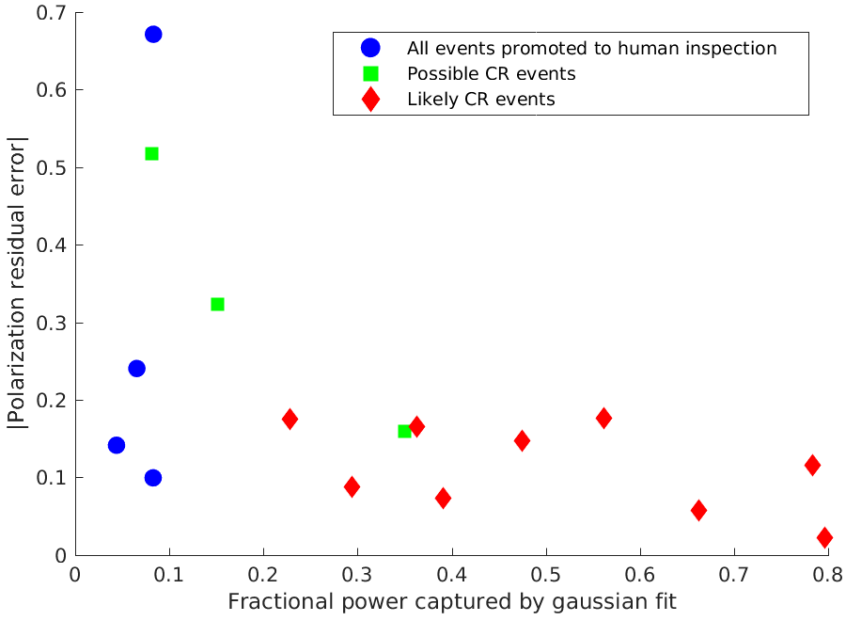


Figure 5.11: **X-axis:** A measure of the fraction of power represented by a spatial Gaussian, as described in Section 5.6. **Y-axis:** A measure of the deviation of polarization signature from the expected one, also as described in Section 5.6. Automated filtering of the "human inspection" dataset performs reasonably well.

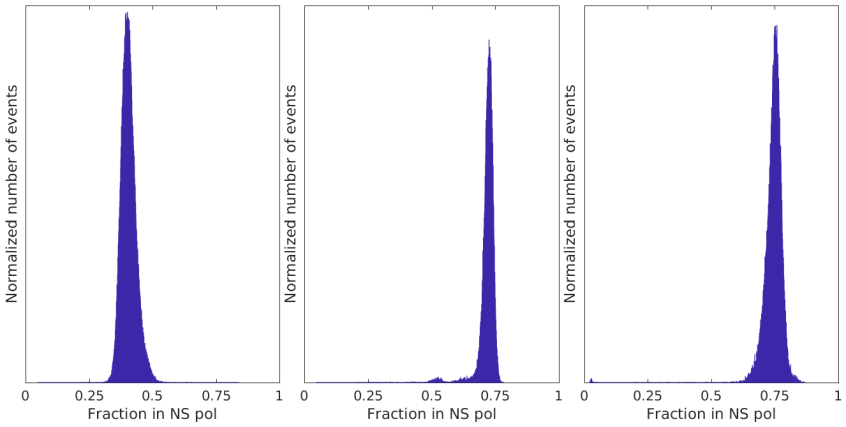


Figure 5.12: A histogram of fractional power received in the N/S-oriented dipole for three distinct RFI events, chosen for high event rates in those azimuth angles. The left two figures represent events originating from different locations within the nearby city of Bishop, CA, while the leftmost plot represents a source coming from somewhere South of the array (there are many candidate sources). Alongside instrumental uncertainties, the noisy parameter fitting of events shown here contributes to the error residuals shown in Figure 5.11

5.7 Characterization of events

A multitude of impulsive events was detected by this system. The more interesting classes of events are listed here, alongside a prototypical event overview for at least

one instance of each type

1. Band-limited impulsive events, coming from the nearby cities of Big Pine and Bishop, or one of another $\sim$ 10 nearby sites of human civilization. These are the overwhelming majority of events detected by the array; Figure 5.13.

2. Events originating from A/C unit on the OVRO-LWA signal processing shelter: Figures 5.14, 5.15, 5.16.

3. Overhead airplanes, many of which traveling to the nearby Eastern Sierra Regional Airport (16 km away with line-of-sight) or Mammoth Yosemite Airport (60 km away). These are characterized by smooth curves in azimuth and elevation against time. It is likely that most sites will experience less RFI from airplanes, provided that they are farther from a very active airport; Figure 5.17.

4. A family of events exhibited all of the properties of cosmic rays, as discussed above; Figure 5.18, 5.19. These events are summarized in Table 5.3. There are nine such events for which the signature is pretty clear, plus another three which are more questionable and therefore discarded in favor of data purity.

5. Ambiguous events: Most events which reach manual inspection are not associated with any other spatial and/or temporal cluster and are simply inspected for a reasonable power distribution across the array (again, polarization was only ever used for validation). However, the challenge remains to discriminate between an isolated RFI event coming from an airplane and one originating from a cosmic ray. This must be validated using the spatial power distribution of the event as seen by the array. Strong events at high elevation angles are easily confirmed, but events coming at low elevation angles can have a very broad power distribution – appearing uniform over the small 200m footprint of the OVRO-LWA core, whereas low power events can have a very weak signature. In both cases, the discrimination can be challenging. The low elevation angle case could be mitigated by using the 1.4 km wide OVRO-LWA expansion array, but being merely a demonstration that the OVRO-LWA is capable of detecting cosmic ray events, the entire issue is sidestepped in this work by discarding all events considered to be "questionable". Future projects could constrain power distribution fits based on the estimated power and direction of arrival, as well as carefully analyze airplane flight patterns and other

backgrounds, allowing for more sensitive discrimination. However, this is beyond the scope of showing the plausibility of OVRO-LWA as a standalone cosmic ray detector and will be reserved for a future project.

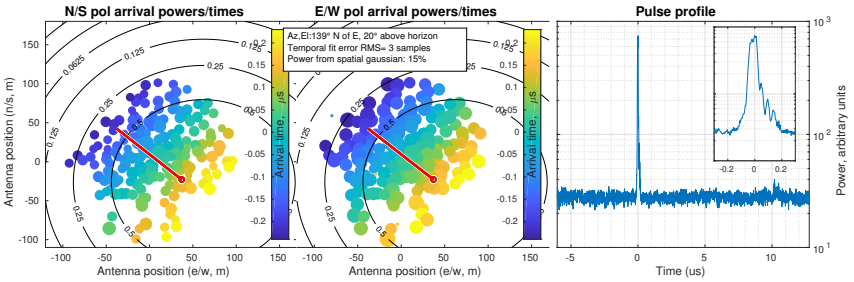


Figure 5.13: Event originating from the nearby city of Bishop, CA. Left, center: Calibrated relative power for each antenna/polarization as size of the scatter-plot circle, time of arrival (in microseconds) in color. The drawn arrow indicates azimuthal direction of travel, but not necessarily location of the strike on the array. Right: Estimated pulse profile constructed through a power sum across all inputs, using the geometric fit extracted from input time of arrivals (pulse frequency content varies somewhat with distance to shower axis, which may partially describe measured pulse profile shape variation).

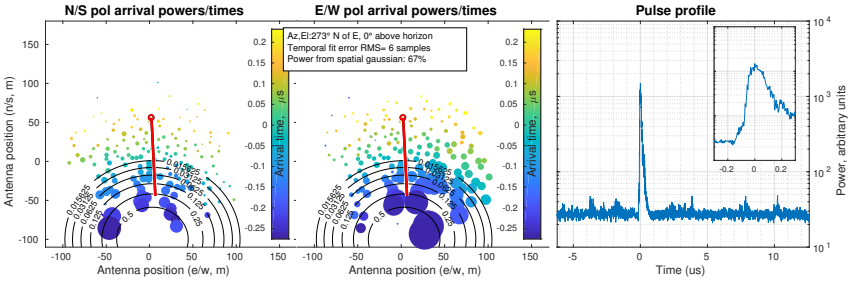


Figure 5.14: "Shelter Event". Wavefront curvature is clearly visible in the left two subplots: events originating from this source appear to be the only ones which are close enough to produce a useful range fit. Note that the pulse profile has extended structure uncharacteristic of a band-limited impulse. Full description as per Figure 5.13.

5.8 Mistakes that were made and other unsorted comments

Since this is one of the first successful RF self-triggers of cosmic rays with a radio telescope, a frank commentary on mistakes and "lessons learned" merits discussion for future efforts.

- **Adjust Linux settings to avoid dropping packets.** A bit of effort was necessary to avoid dropping packets at high dump rates: it is important to

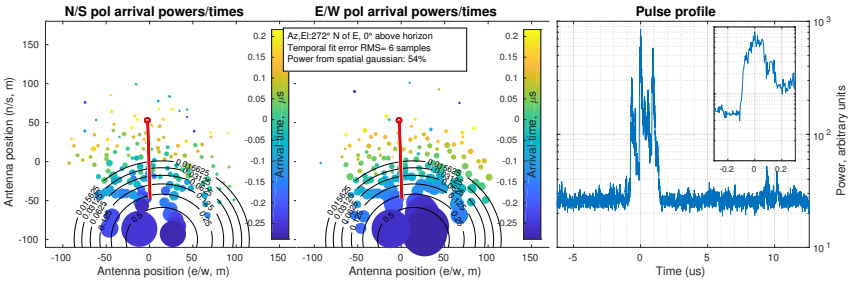


Figure 5.15: Another event originating from A/C unit on OVRO-LWA signal processing shelter. Note the multiple distinct components. It is believed that this is caused by a relay toggling as the A/C unit is activated, judging by past experience as well as a temporal pattern of events which is synchronous with A/C activity. Full description as per Figure 5.13.

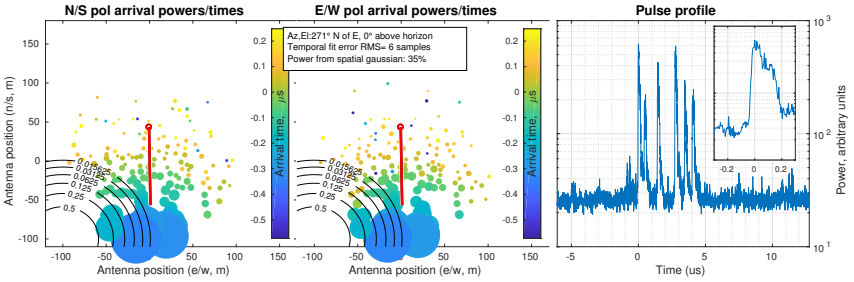


Figure 5.16: Third event likely originating from A/C unit on OVRO-LWA signal processing shelter. Despite the multiple peaks in the pulse profile which confused an early stage of the algorithm, robust plane fitting of the time-of-arrivals allowed successful location fitting. Full description as per Figure 5.13.

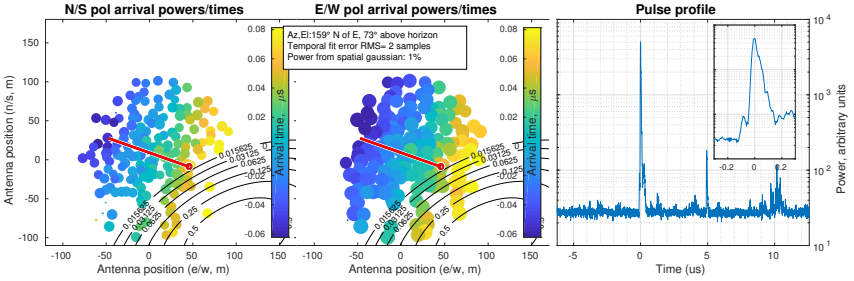


Figure 5.17: Non-cosmic ray event originating from above the array - presumably an airplane, considering other spatio-temporally correlated events. Most detected events originating from airplanes had a single component. The event is unlikely to be caused by reflections from nearby RFI sources because the line-of-sight event would have been brighter and triggered the system first. Full description as per Figure 5.13.

limit maximum filesize in tcpdump, and direct tcpdump to save to the Linux machines' ramdisk – failure to do so resulted in substantial packet loss at

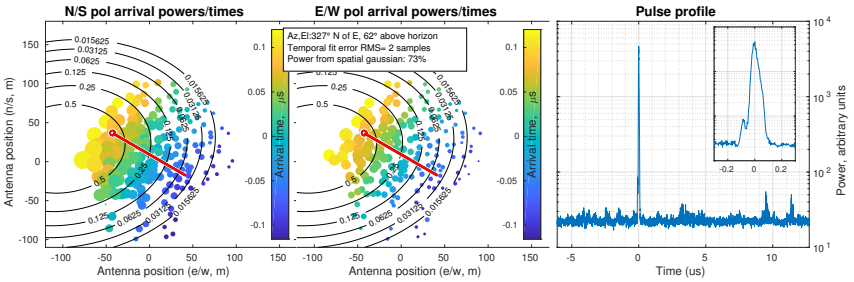


Figure 5.18: Event likely to originate from a cosmic ray.

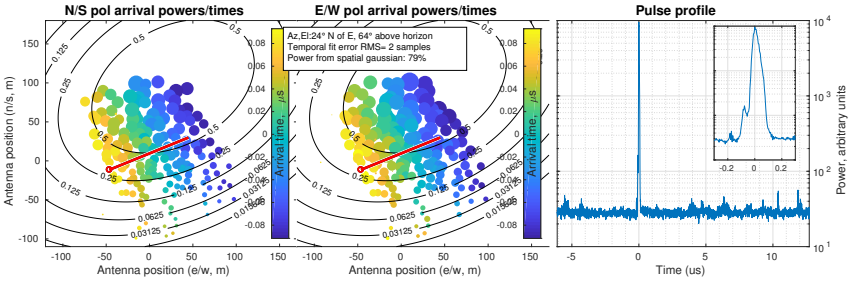


Figure 5.19: Event likely to originate from a cosmic ray.

$\sim$ 20 events/sec, versus $\sim$ 300 events/sec. Use another process to move files from the ramdisk to more permanent storage. Likewise, default UDP receive buffers must be increased from their (small) defaults each time the machine is booted: the commands "sysctl -w net.core.rmem_max=26214400; sysctl -w net.core.rmem_default=26214400" were used for this purpose. On a related note, UDP packet fragmentation was unavoidable given our network configuration, and had to be corrected in post-processing. Since each packet was split into three fragments, dropping of any of those fragments resulted in the loss of all data for that input.

- **Cross-correlation time-of-arrival method** The initial time-of-arrival algorithm was far more complicated than the one presented in this work, including estimation of the received signal profile with almost no priors and then using the peak cross-correlation between that estimated signal and all inputs as the time of arrival estimate. It was found that almost all RFI events detected by this system had a strong impulsive component (not surprising, considering that was what the FPGA detector was designed to operate on). The technique described in Section 5.6 is faster and can give useful results at lower SNR, but exhibits larger fitting residuals than the more sophisticated technique. That said, the Hilbert envelope technique described in (Schellart, A. Nelles, et al.,

Azimuth	Elevation	Power	% seen→N/S	% expec→N/S	Error	Class
327.3°	61.7°	177.8	34	22	-12	CR
96.7°	47.9°	17.7	84	99	15	CR
94.7°	40.3°	55.1	92	99	7	CR
24.4°	63.8°	361.8	55	57	2	CR
206.1°	42.4°	30.8	9	4	-6	CR
72.5°	63.2°	20.6	84	93	9	CR
164.4°	40.0°	10.7	54	38	-17	CR
172.5°	44.3°	12.8	54	37	-18	CR
327.7°	80.2°	11.9	65	75	10	CR
24.6°	48.8°	64.2	61	43	-18	CR
85.0°	20.2°	21.3	67	100	32	NC
202.3°	33.8°	24.5	53	1	-52	NC
10.1°	54.9°	15.2	54	38	-16	NC
84.0°	20.6°	10.1	32	100	67	RFI
204.6°	11.5°	19.2	24	0	-24	RFI
201.3°	5.9°	20.2	16	2	-14	RFI
201.2°	7.0°	20.3	16	2	-14	RFI
206.3°	42.2°	31.7	09	3	-6	RFI

Table 5.3: Cosmic ray candidates, grouped by classification (CR: Cosmic Ray, NC: No Call, RFI: Radio Frequency Interference), subsequently sorted in order of arrival time. Power is measured in multiples of the galactic background power (estimated as the pulse profile estimate far before event arrival:median across all events in the dataset, in spite of the fact that galactic background power changes as a function of local sidereal time). Fractional power received each polarization estimated as the amplitude of a fit spatial Gaussian in a Gaussian+DC model. A rudimentary beam model was used for the expected power for each polarization, but that beam model has only been weakly validated and is subject to several instrumental effects.

2013) would have likely have performed very well as well. It is likely that a cross-correlation between input voltages would have worked best of all (provided that a full cross-correlation of all inputs was performed). This was prototyped and worked well, but the best way to extract the relative time-of-arrival for each input was not obvious in the context of not all inputs receiving an impulse, resulting in the use of the techniques described above. If a voltage cross-correlation technique was used however, precise extraction of individual antenna power estimates could be performed by formulating the problem as a noisy Rank-1 matrix completion problem, where each element in the matrix is the peak cross-correlation power between two inputs, and the diagonal (to be solved for) is the power seen by each input. Such a technique would largely remove the additive impact of receiver noise on power estimates, leaving only

the astronomical background (this would likely give performance similar to imaging under the assumption that the image consists of only thermal noise and a point source). If cross-correlating voltages, remember to correlate the analytic version of all signals (via a Hilbert transform, for example).

- **Distribution of error sources** The RMS time-of-arrival residual was about 4 samples with the arg-max time-of-arrival estimator, and about 3 samples with the more cross-correlation technique (available in the included code, but not described here for brevity). In each case, about 0.5 samples of that residual was due to using the arg-max of the signal power (or alternatively, cross-correlation value), which does not consider the value of nearby samples. A better technique would use some form of super-resolution, such as quadratic interpolation of the peak location, to generate sub-sample resolution time-of-arrival estimates. An additional ~ 1.8 samples of the RMS residual was due to bias that was systematic to each antenna – likely inaccuracies in either cable delay estimation or antenna position estimation. Accurate estimation of these parameters would also have improved system performance. Responding to this issue, cable-length delays for each input were estimated using triggered RFI events. Great care was assigned to avoiding biasing the delay estimates towards common events. If this work were repeated, the same process would be performed jointly on antenna positions as well. It is important to note that with many FPGA boards and ADCs, the on-FPGA PLLs will lock to a different clock cycle of the ADC data line on each FPGA configuration. For this reason, it is possible that the effective cable lengths will vary across FPGA configurations, requiring a new cable length calibration each time.

- **You want the original samples** Being able to process the raw ADC voltages directly (as opposed to F-engine products, which lack the needed temporal resolution, or beam-formed products which lack the needed spatial coverage) was crucial for the success of this project. Additionally, being able to program the FPGAs with custom firmware greatly simplified the effort.

- **Particle detectors would have helped.** Having on-site particle detectors for the first several detections would have made establishing priors on the cosmic ray signal for this instrument much easier, thereby accelerating development.

- **Power line RFI is 60 Hz phase dependent.** Although it was not demonstrated in this work, many sources of terrestrial RFI are correlated with the phase

of the 50/60 Hz mains voltage: filtering event triggers on this signal may produce good RFI mitigation results for some sources. Because λ at 60 Hz is ~50,000 km, phase acquisition could be done at a central location and then transmitted to a distributed processing system, provided adequate time synchronization.

- **Matched filtering would have done better.** Matched filters are the optimal detector of a known signal in the presence of white Gaussian noise (Turin, 1960). This would likely have performed better than a moving average. Depending on the bandwidth of the instrument and RFI, it is possible that the impulse response of the system can be used as this matched filter (for OVRO-LWA, this was the case).

- **Grouping of antennas.** All detected cosmic rays, as well as almost all RFI sources, were considerably polarized. If grouping antennas on FPGAs for detection purposes, it might have been helpful to address polarization directly by taking the time-of-arrival for the highest SNR input on each antenna or alternatively simply summing in power across polarizations for the purpose of detection and localization. Likewise, re-arranging antennas on FPGAs such that each FPGA only contained inputs corresponding to one polarization could be an alternative method to capitalize on this property. Doing this would give the direction-of-arrival algorithm more unique locations with which to fit and presumably greater spatial extent. Since most of the power was typically in one polarization or the other for this work, sorting by polarization would not have terrible consequences on sensitivity for most RFI events, most of the time. None of this was explored, but it is worth thinking about the trade-offs of various antenna grouping schemes for future experiments.

- **Elevation filtering was not as important as expected**. Initial filters rejected all events with an elevation angle less than 29°, among many other filters which were tested at one point or another in this work. As an experiment to see if the system as designed would be effective at detecting τ-neutrinos, this filter was relaxed to permit all events above an elevation angle of 2°. This resulted in more false positives (by a factor of $2 \sim 3$) being promoted to human inspection, but no additional missed detections (which was a small risk due to the time domain clustering algorithm). Adding more filter metrics (reject events which lack data from over 1/4th of inputs due to instrumental errors; reject events whose pulse profiles do not appear to represent band-limited

impulses) brought the RFI rejection performance to that which is shown in the main body of this work. Because the algorithm works well down to $\sim 2^\circ$ elevation angle, it is likely suitable for discriminating between RFI and τ-neutrinos.

- **The 3D source fitting algorithm performed much better than plane fitting but was computationally expensive.** The 3D algorithm was written in MATLAB and likely very computationally inefficient. Likewise, the faster but less accurate robust plane-fitting algorithm was simply called from MATLAB without any special tuning. It is possible that engineering could bridge the gap between these algorithms. Likewise, implementing the entire routine in a compiled language (as opposed to MATLAB) could render the entire issue irrelevant.

- **These automatic RFI filtering metrics completely ignore the distribution of antenna power as a function of position.** We held the opinion that it would be too challenging to algorithmically discriminate RFI from astronomical events without having examples of cosmic rays. Data from both RFI events and cosmic rays is available to assist future experimenters in designing triggers. The full summarized dataset can additionally be provided upon request (80 GB). It is our opinion that designing this filter for cosmic rays is practical given the available data. The first draft of such a filter is presented in Figure 5.11 and works reasonably well.

- **There was an unexpected surplus of RFI at** 15° **elevation angle**. We put a fair amount of effort into solving this, and could not. Although it has not been published, private communication indicates that the AREA team has also seen similar behavior, which is also visible (but not addressed) in Figure 2 of (Kelley, 2013). Through inspection of the frequency of narrow-band RFI and comparison with imaging data and spectra, we are confident that the sampling frequency is correct. Several sources known to be produced on the ground near the array are reconstructed to positions near the horizon. There are no locations near OVRO-LWA which are perceived as a 15° elevation angle from the location of the array. However, hand-calculation of the elevation angle of these strange events results in solutions very similar to those produced in the automated fashion. Reflections against some subterranean surface were considered but rejected based on signal strength arguments. Figure 5.20 shows the histogram of observed elevation angles in detail, while the upper-

left section of Figure 5.10 illustrates the behavior in a different perspective. The issue remains a mystery.

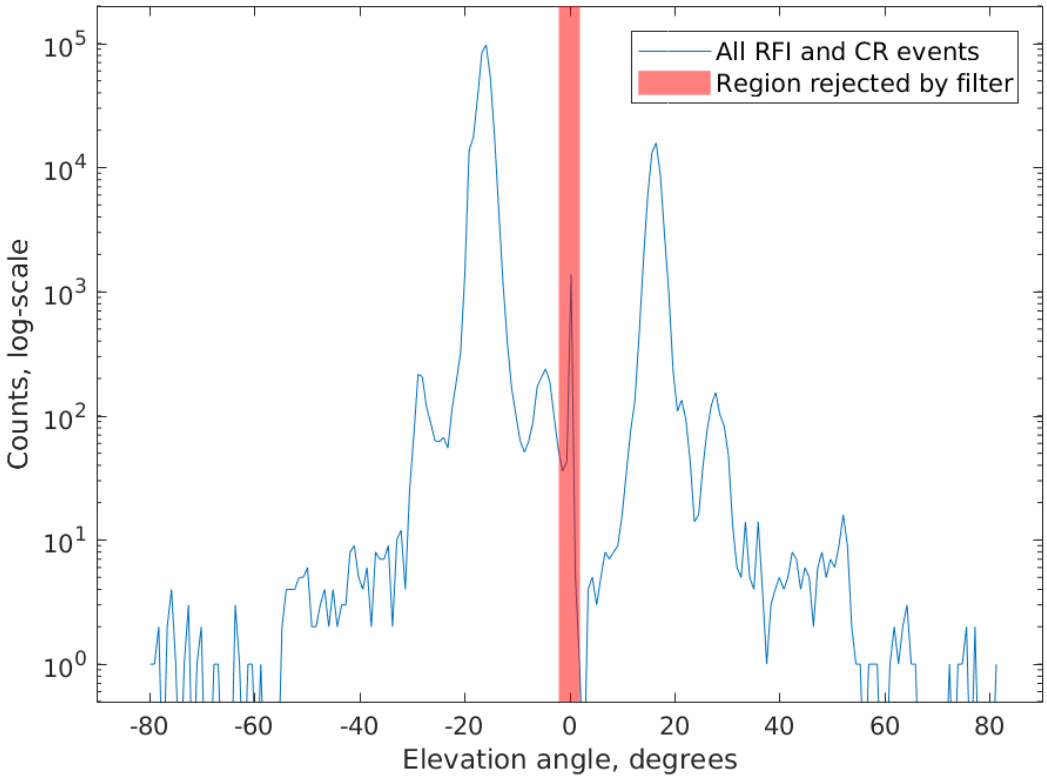


Figure 5.20: Histogram of event elevation angles. The array is nearly planar, producing a degeneracy between events above the array and below the array. For all other elevation angle displays in this work, the absolute value of the elevation angle was taken.

5.9 Future experimental design

The lessons learned from this effort (and prior work in the field) inform future experiments. Broadly speaking the method of using single FPGAs for batch-processing baseband sampled antennas as described in Section 5.3 was good. Future systems should perform techniques similar to those described in Section 5.5, as well as (5.11), (5.12), and (5.13) on-FPGA. Afterward, the subsequent steps of source direction fitting with a plane-wave assumption should be performed on-board (Josh and Kleeman, 2016) (FPGA soft processors could also be used for this task), followed by azimuth/elevation/fit-quality RFI blocking, which is sent to other devices as a time-stamped message to other FPGAs, not unlike the trigger packets used for regular detection above (Figure 5.21). This would filter all except ~1/10,000 events with minimal consequence to cosmic ray detections, suitable for sites with relatively

little air traffic. At these "low air traffic" sites, the data pipeline will simply saturate while the airplane is overhead, likely preventing cosmic ray detections. For sites with heavier air traffic or severe bandwidth limitations (such as those communicating event detections over cellular), these "Stage 2" detections can instead be buffered in on-board DRAM. After several minutes have passed, either time-domain clustering or robust curve fitting on {azimuth, elevation, time-stamp} as described in Section 5.8 can be used to flag out airborne RFI sources, reducing event rate to ∼1/hr if applied at the OVRO-LWA, which is slow enough for most arrays.

In order to simulate the impact of an algorithm which uses on-chip plane-fitting algorithms such as those described above, previously collected data was re-processed. Antennas were re-clustered into groups of 16 which were more optimal for RFI filtering, and source localization was performed using robust plane fitting, with the omission of the better-performing 3D localization routines. A separate plane fit was performed with each of the 16 groups of antennas, with the lowest-residual fit being accepted as the "true" source direction. The algorithm performed reasonably well – worse than a full fit, but would be a much more effective RFI filter than the methods described in Section 5.5, especially considering that PC post-processing can easily flag the remaining false positives once all data is available. A quick demonstration of the feasibility of this technique is provided in Figure 5.22, which shows similar behavior and quality when performing a simplified direction of arrival routine on-chip.

A successful RF self-trigger system relies on having (on a processing unit scale) both a sufficient density of antennas (to have coincident detections on cosmic rays) and sufficient extent of antennas (for the direction of arrival estimation). For the system and data analyzed in this work, a ∼100 m maximum antenna separation would have been suitable for source localization on these high elevation angle events. For arrays which collect all signals at a central location, this suggests arranging inputs on FPGAs such that they have at least roughly this extent in both dimensions. The narrow E/W extent of the OVRO-LWA antenna-FPGA groupings appears to have made it less suitable for this style of filtering.

For arrays which only collect post-trigger data, those same processors could be placed in the field, handling small groups of antennas. The same processing for centralized arrays could then be performed by sending the {azimuth, elevation, time-stamp} data over cellular to a central server which, after rejecting all airplanes, would trigger a dump from the local DRAM buffers to the central server (Figure 5.23).

Alternatively, beamforming the array in various directions would make near-optimal use of the collecting area of the array. Doing this right is tricky because you cannot simply beamform the entire array in each direction: the event may strike only a subset of the array antennas. Therefore, an intelligent trade-off will have to be struck between spatial resolution and detection SNR, optimized for the expected CR footprint on the array. After an initial detection is made in some direction, beamforming groups of antennas in the array in the direction implied by wavefront fitting will maximize SNR for CR parameter extraction. Using hierarchal beamforming (Zarb-Adami et al., 2010; Hampson, de Wild, and Smolders, 2000) will make the computation of beams economical with minimal SNR cost.

Plans at OVRO include near-term expansion to 352 antennas, as well as expansion to 2048 dipoles on a timescale of ~5 years. Both arrays will span a 2.6 km diameter area and will include extensive air-shower detection as a continuous commensal of operation. Using the lessons learned from this work, the techniques described in this section, as well as Section 5.8, will be applied directly in the design of that cosmic ray engine.

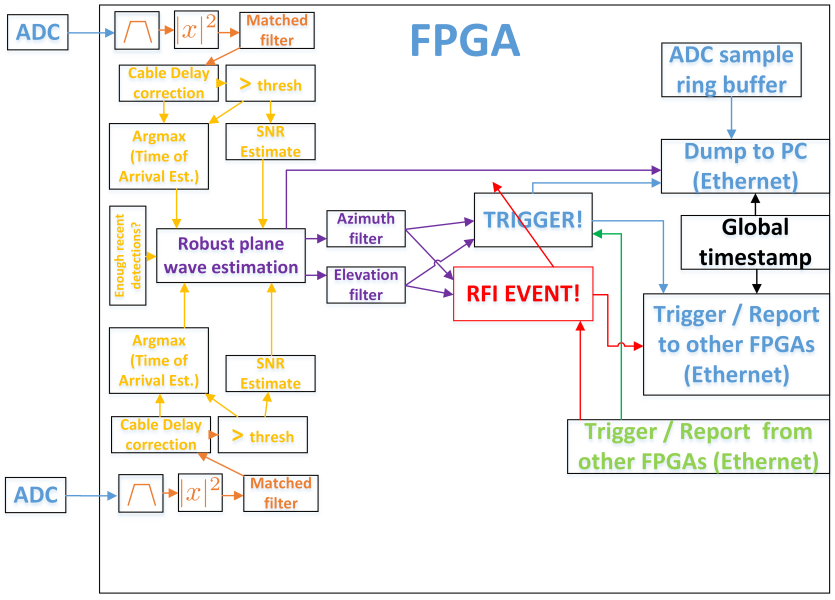


Figure 5.21: Block diagram of a proposed, lower bandwidth FPGA firmware. In prototypes, robust planar fitting vastly outperforms the RFI filtering used in Section 5.5.

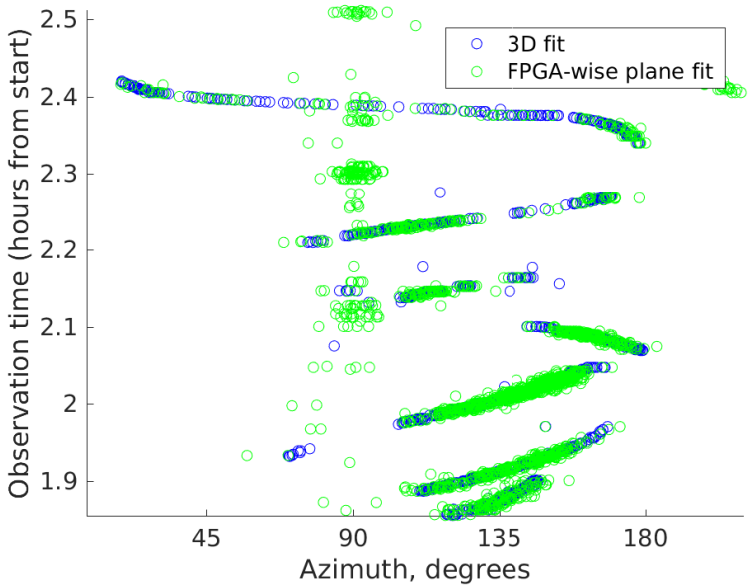


Figure 5.22: Similar analysis to that performed in Section 5.6, overlaid on top of the results of local planar fits from individual FPGAs as described in Section 5.9. Software issues resulted in the spurious events around Azimuth 90°, which would not be present in a production system.

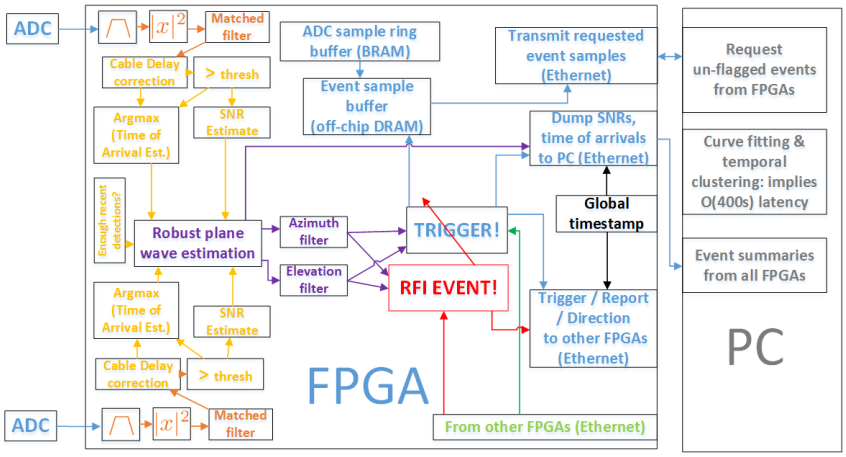


Figure 5.23: Block diagram of a proposed, lower bandwidth FPGA firmware, with the use of an additional PC co-processor to further reduce bandwidth by avoiding the transmission of airplane-driven or temporally clustered events.

Beam-mapping of radio-astronomy arrays using cosmic ray events

Understanding the individual antenna response of low frequency radio-astronomy arrays can be very challenging. For higher frequency telescopes, steering the dish across a bright point source allows for the response to be mapped. However, low

frequency antennas are not large compared to the wavelength of radiation that they receive, and so individual antennas/receivers typically see a substantial fraction of the sky at any given time. Additionally, these antennas are fixed and cannot be steered beyond what the motion of the earth facilitates. Techniques to reduce the sky to a single point source include using man-made drones as artificial point sources or taking advantage of the variability of radio pulsars to make a time-domain filter which rejects the rest of the sky. The drone technique suffers from challenges mapping the antenna pattern of the drone, and current techniques only map the amplitude of each antennas beam. On a similar note, pulsar holography suffers from a limited selection of pulsars which have adequate SNR (limiting the region of the beam which can be mapped). Cosmic ray events provide an alternate technique by which the antenna beams can be mapped. Each event listed in this work has an individual antenna SNR of greater than five for a substantial portion of the array, and several events were ten times that bright. First, a cosmic ray event would be detected and its parameters (such as the direction of arrival, energy and X_{max}) extracted. A model of the complex-valued gain seen by each antenna (to a scalar power constant and a scalar time of arrival, likely as a function of frequency) is extracted. This is compared to the actual complex-valued gain seen by each antenna. This allows the relative gains of antennas to be mapped. The technique described here does not provide an absolute gain in each direction, but does greatly simplify the process of mapping individual dipole beams, provided a sufficient density of cosmic ray events striking the array. Making this work in practice will be challenging–the power received (and therefore resultant SNR) will vary as a function of event and antenna. Cosmic ray parameter estimation errors (including the direction of arrival) will contribute to errors in beam modeling, especially for cosmic rays with low received power. There will be a frequency-resolution/SNR trade-off. One good approach might be to grid events spatially. On an event-by-event basis (again, as a function of frequency, per an FFT), remove delays implied by geometric effects and a hyperbolic or conical beaming model at a sub-sample resolution. Integrate the expected and actual signal received after removing a model of thermal noise. After integrating a sufficient number of events, take the complex-valued ratio between the expected and actual received signal, as a function of antenna, frequency, and direction. That value will be an estimate of the complex-valued relative gains of each antenna in the given direction. On a related note, this same technique could be performed on the ~ 5000 airplane events which were detected in this dataset (with the added challenge that the source signal in those cases is unknown). It is possible that these airplane-driven

events are not isotropic (A. Nelles, 2018), but AREA (the only other group known to investigate this in any capacity–it appears that they were trying to synchronize their antennas with better accuracy) only saw one useful airplane per week(Aab and al., 2016; T. Huege, 2016b), whereas OVRO-LWA appears to see dozens per day. This warrants further investigation.

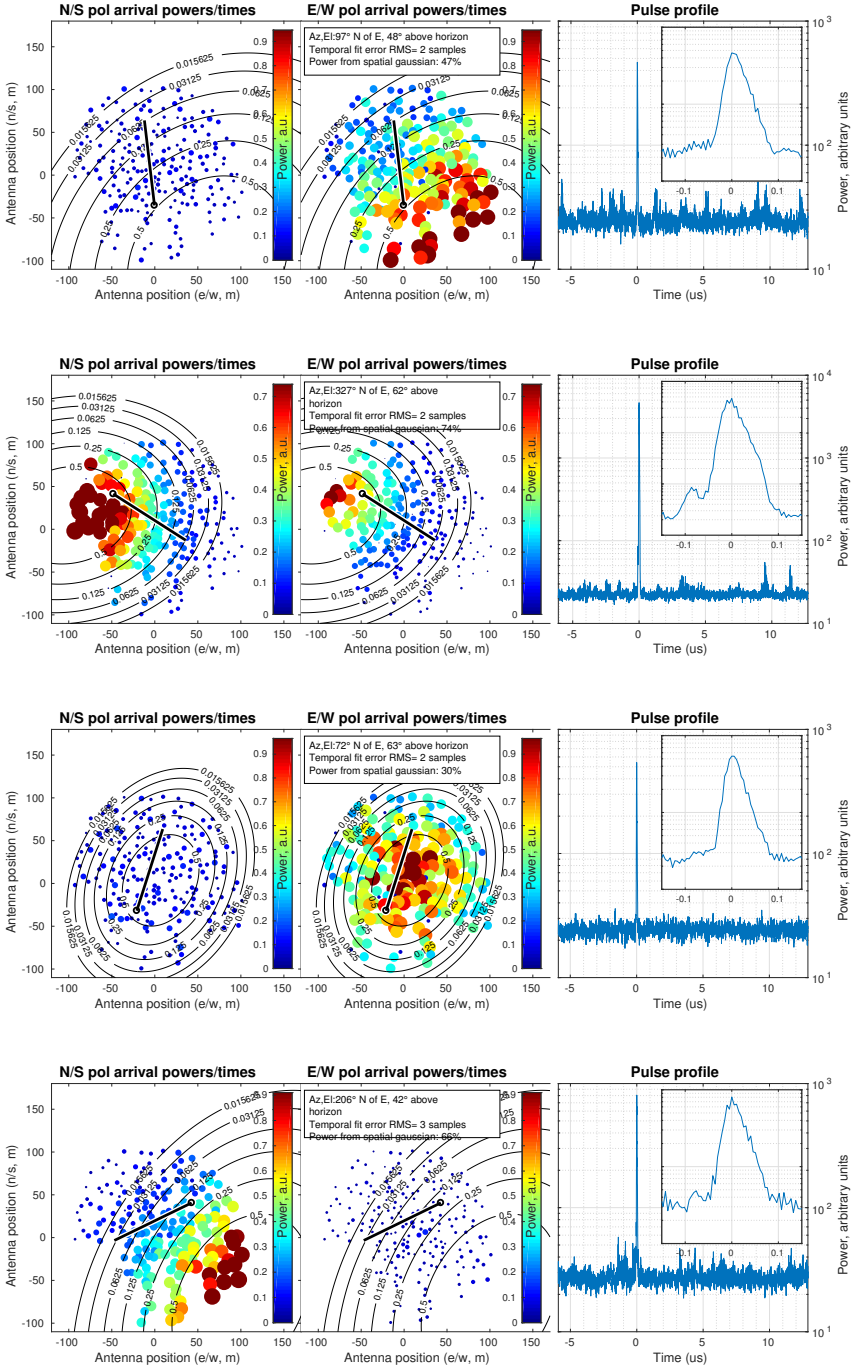
5.10 Conclusion

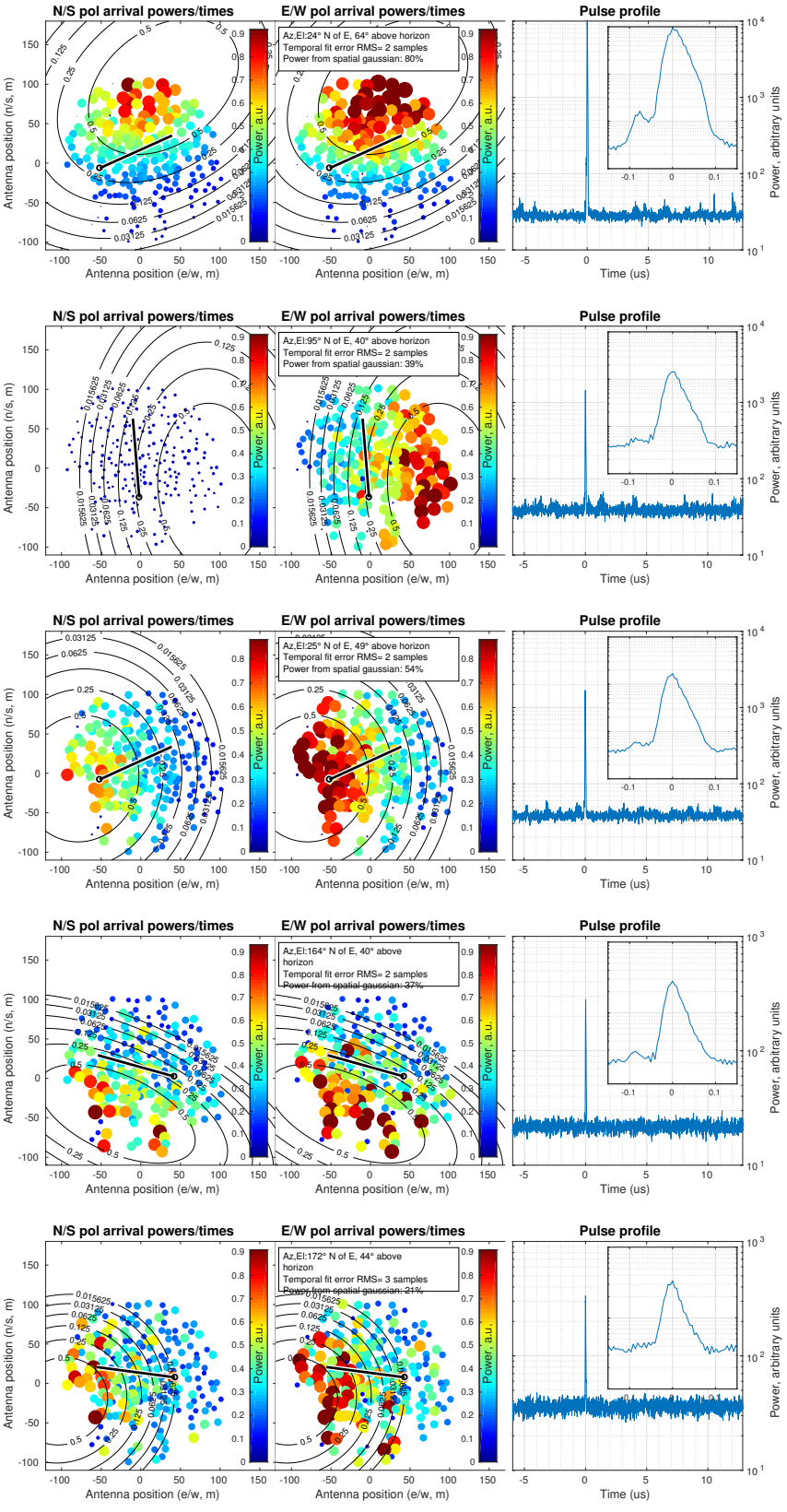
Purely RF detection of cosmic rays has been demonstrated with the OVRO-LWA using custom FPGA firmware and software post-processing. The RFI environment was a challenging factor which was surpassed with effective temporal and spatial filtering on both FPGA and PC. This demonstration, although it can improve, lends credibility to future arrays of RF-only air shower detectors for cosmic ray and neutrino science. All software and firmware, as well as a subset of the raw data which includes cosmic ray detections, is available (Monroe, 2018). Look forward to the final expansion of the OVRO-LWA, with a full duty-cycle, real-time cosmic ray engine!

5.11 All events promoted to human inspection

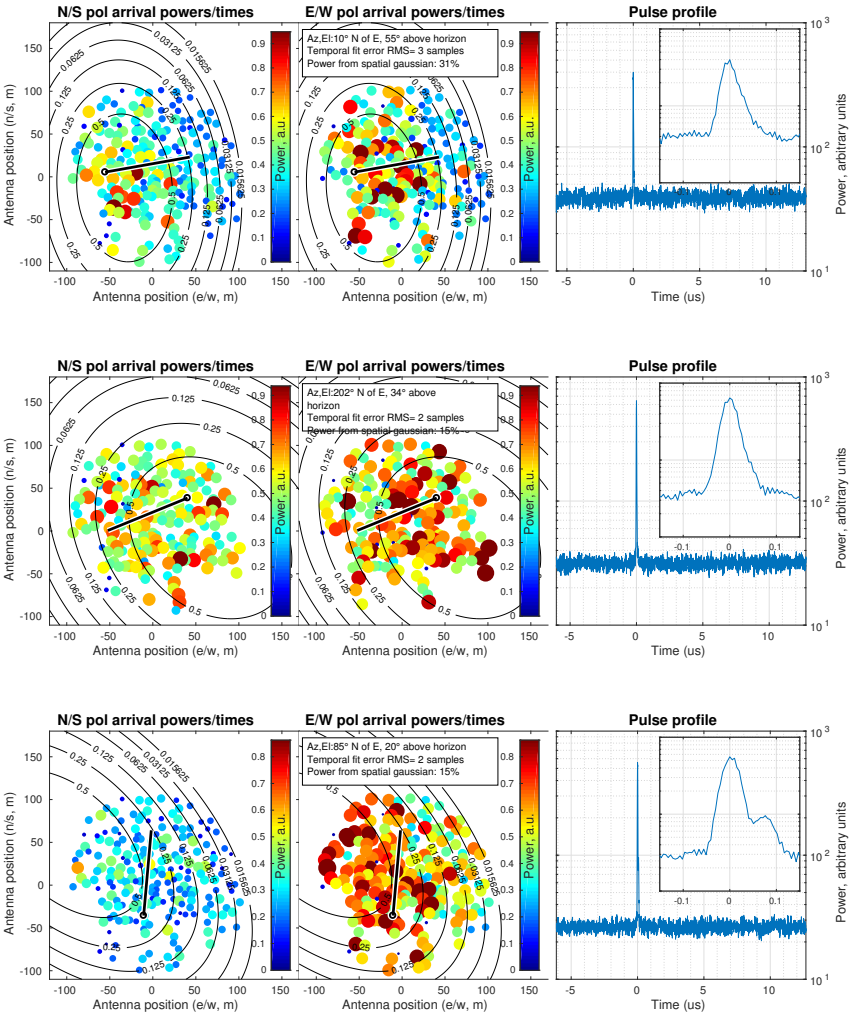
In order to better visualize the spatial power distribution of the events, these plots represent the power received as a function of antenna position in both the color *and* area of each scatter-point, a departure from previous plots which contain a similar aesthetic.

Events likely to originate from cosmic rays

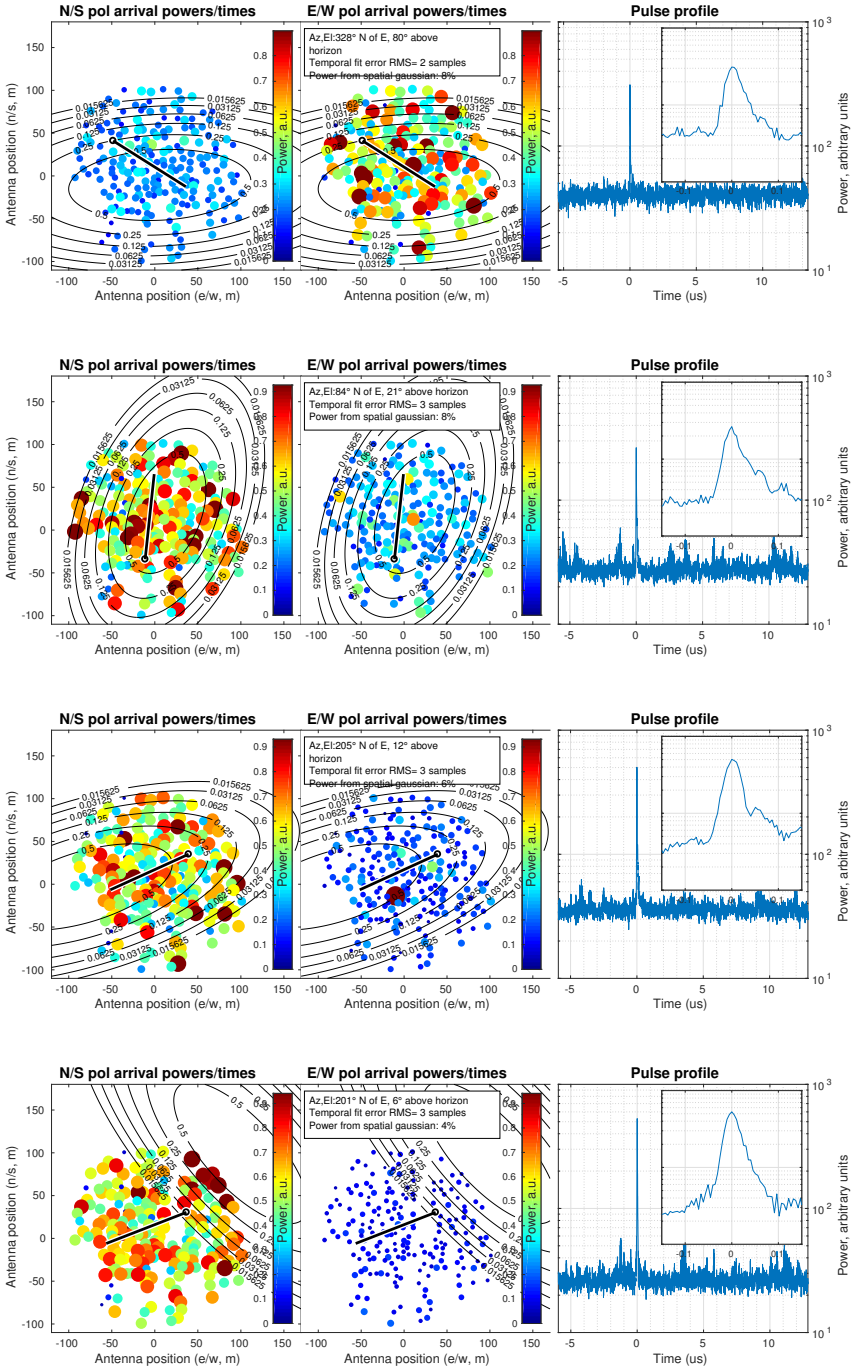



N/S pol arrival powers/times
E/W pol arrival powers/times
Pulse profile
Az,El:24° N of E, 64° above horizon
Temporal fit error RMS= 2 samples
Power from spatial gaussian: 80%
Az,El:95° N of E, 40° above horizon
Temporal fit error RMS= 2 samples
Power from spatial gaussian: 39%
Az,El:25° N of E, 49° above horizon
Temporal fit error RMS= 2 samples
Power from spatial gaussian: 54%
Az,El:164° N of E, 40° above horizon
Temporal fit error RMS= 2 samples
Power from spatial gaussian: 37%
Az,El:172° N of E, 44° above horizon
Temporal fit error RMS= 3 samples
Power from spatial gaussian: 21%
Antenna position (n/s, m)
Antenna position (e/w, m)
Power, a.u.
Time (us)
Power, arbitrary units

Events which are inconclusive



Events likely to originate from man-made sources



References

Aab, A. and et al. (2016). “Nanosecond-level time synchronization of autonomous radio detector stations for extensive air showers”. In: *Journal of Instrumentation* 11.01, P01018. URL: http://stacks.iop.org/1748-0221/11/i=01/a=P01018.

Aab, A. et al. (2016a). "Energy estimation of cosmic rays with the Engineering Radio Array of the Pierre Auger Observatory". In: *Physical Review D* 93.12, 122005, p. 122005. DOI: 10.1103/PhysRevD.93.122005. arXiv: 1508.04267 [astro-ph.HE].

– (2016b). "Measurement of the Radiation Energy in the Radio Signal of Extensive Air Showers as a Universal Estimator of Cosmic-Ray Energy". In: *Physical Review Letters* 116.24, 241101, p. 241101. DOI: 10.1103/PhysRevLett.116.241101. arXiv: 1605.02564 [astro-ph.HE].

Aartsen, M. G., R. Abbasi, et al. (2013). "Measurement of the cosmic ray energy spectrum with IceTop-73". In: *Physical Review D* 88.4, 042004, p. 042004. DOI: 10.1103/PhysRevD.88.042004. arXiv: 1307.3795 [astro-ph.HE].

Aartsen, M. G., K. Abraham, et al. (2016). "Constraints on Ultrahigh-Energy Cosmic-Ray Sources from a Search for Neutrinos above 10 PeV with IceCube". In: *Physical Review Letters* 117.24, 241101, p. 241101. DOI: 10.1103/PhysRevLett.117.241101. arXiv: 1607.05886 [astro-ph.HE].

Abraham, J. et al. (2008). "Upper Limit on the Diffuse Flux of Ultrahigh Energy Tau Neutrinos from the Pierre Auger Observatory". In: *Physical Review Letters* 100.21, 211101, p. 211101. DOI: 10.1103/PhysRevLett.100.211101. arXiv: 0712.1909.

Abreu, P. et al. (2012). "Results of a self-triggered prototype system for radio-detection of extensive air showers at the Pierre Auger Observatory". In: *Journal of Instrumentation* 7, P11023. DOI: 10.1088/1748-0221/7/11/P11023. arXiv: 1211.0572 [astro-ph.HE].

Anderson, M. M. et al. (2017). "A simultaneous search for prompt radio emission associated with the short GRB 170112A using the all-sky imaging capability of the OVRO-LWA". In: *ArXiv e-prints*. arXiv: 1711.06665 [astro-ph.HE].

Ardouin, D. et al. (2009). "Geomagnetic origin of the radio emission from cosmic ray induced air showers observed by CODALEMA". In: *Astroparticle Physics* 31, pp. 192–200. DOI: 10.1016/j.astropartphys.2009.01.001. arXiv: 0901.4502 [astro-ph.HE].

ASTRON (2018). *DATA QUALITY INSPECTION*. URL: https://www.astron.nl/radio-observatory/observing-capabilities/depth-technical-information/data-quality-inspection/data-qu (visited on 03/07/2018).

Barwick, S. W. et al. (2017). "Radio detection of air showers with the ARIANNA experiment on the Ross Ice Shelf". In: *Astroparticle Physics* 90, pp. 50–68. DOI: 10.1016/j.astropartphys.2017.02.003. arXiv: 1612.04473 [astro-ph.IM].

Bezyazeekov, P. A. et al. (2015). "Measurement of cosmic-ray air showers with the Tunka Radio Extension (Tunka-Rex)". In: *Nuclear Instruments and Methods in Physics Research A* 802, pp. 89–96. DOI: 10.1016/j.nima.2015.08.061. arXiv: 1509.08624 [astro-ph.IM].

Buitink, S. and et al. (2016). "A large light-mass component of cosmic rays at 10^{17} – $10^{17.5}$ electronvolts from radio observations". In: *Nature* 531, 70?73.

Buitink, S., A. Corstanje, et al. (2014). "Method for high precision reconstruction of air shower X_{max} using two-dimensional radio intensity profiles". In: *Physical Review D* 90.8, 082003, p. 082003. DOI: 10.1103/PhysRevD.90.082003. arXiv: 1408.7001 [astro-ph.IM].

Corstanje, A. et al. (2011). "LOFAR: Detecting Cosmic Rays with a Radio Telescope". In: *ArXiv e-prints*. arXiv: 1109.5805 [astro-ph.HE].

Devices, Analog (2015). *HMCAD1511*. http://www.analog.com/media/en/technical-documentation/data-sheets/hmcad1511.pdf.

Dexter, Matt and et al. (2014). *ADC16x250-8 RJ45*. https://casper.berkeley.edu/wiki/ADC16x250-8.

Eastwood, M. W. et al. (2017). "The Radio Sky at Meter Wavelengths: m-Mode Analysis Imaging with the Owens Valley Long Wavelength Array". In: *ArXiv e-prints*. arXiv: 1711.00466 [astro-ph.IM].

Greisen, K. (1966). "End to the Cosmic-Ray Spectrum?" In: *Physical Review Letters* 16, pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.

Hampson, G. A., R. de Wild, and A. B. Smolders (2000). "Efficient Multi-Beaming for the Next Generation of Radio Telescopes". In: *Perspectives on Radio Astronomy: Technologies for Large Antenna Arrays*. Ed. by A. B. Smolders and M. P. van Haarlem, p. 265.

Hoover, S. et al. (2010). "Observation of Ultrahigh-Energy Cosmic Rays with the ANITA Balloon-Borne Radio Interferometer". In: *Physical Review Letters* 105.15, 151101, p. 151101. DOI: 10.1103/PhysRevLett.105.151101. arXiv: 1005.0035 [astro-ph.HE].

Huege, T. (2016a). "Radio detection of cosmic ray air showers in the digital era". In: *Physics Reports* 620, pp. 1–52. DOI: 10.1016/j.physrep.2016.02.001. arXiv: 1601.07426 [astro-ph.IM].

Huege, Tim (2016b). *Nanosecond-level time synchronization of AERA using a beacon reference transmitter and commercial airplanes*. [Online; accessed 30-May-2018]. URL: https://indico.cern.ch/event/461474/contributions/2152679/attachments/1289740/1920214/huege-aera-timingcalibration_v1.pdf.

Inc., Mathworks (2015a). *MATLAB fminunc*. https://www.mathworks.com/help/optim/ug/fminunc.html.

– (2015b). *MATLAB pcfitplane*. https://www.mathworks.com/help/vision/ref/pcfitplane.html.

Josh, Horace and Lindsay Kleeman (2016). "A novel hardware plane fitting implementation and applications for bionic vision". In: *Machine Vision and Applications* 27.7, pp. 967–982. DOI: 10.1007/s00138-016-0764-8. URL: https://doi.org/10.1007/s00138-016-0764-8.

Kambeitz, O. (2017). "Measurement of horizontal air showers with the Auger Engineering Radio Array". In: *European Physical Journal Web of Conferences*. Vol. 135. European Physical Journal Web of Conferences, p. 01015. DOI: 10.1051/epjconf/201713501015. arXiv: 1609.05456 [astro-ph.IM].

Kelley, J. L. (2013). "Data acquisition, triggering, and filtering at the Auger Engineering Radio Array". In: *Nuclear Instruments and Methods in Physics Research A* 725, pp. 133–136. DOI: 10.1016/j.nima.2012.11.153. arXiv: 1205.2104 [astro-ph.IM].

Kocz, J., L. J. Greenhill, B. R. Barsdell, G. Bernardi, et al. (2014). "a Scalable Hybrid Fpga/gpu FX Correlator". In: *Journal of Astronomical Instrumentation* 3, 1450002-330, pp. 1450002–330. DOI: 10.1142/S2251171714500020. arXiv: 1401.8288 [astro-ph.IM].

Kocz, J., L. J. Greenhill, B. R. Barsdell, D. Price, et al. (2015). "Digital Signal Processing Using Stream High Performance Computing: A 512-Input Broadband Correlator for Radio Astronomy". In: *Journal of Astronomical Instrumentation* 4, 1550003, p. 1550003. DOI: 10.1142/S2251171715500038. arXiv: 1411.3751 [astro-ph.IM].

Liu, T. C. et al. (2017). "The status of the second station of Taiwan Astroparticle Radiowave Observatory for Geo-synchrotron Emissions (TAROGE-II)". In: *International Cosmic Ray Conference* 35, 234, p. 234.

Martineau-Huynh, O. et al. (2017). "The Giant Radio Array for Neutrino Detection". In: *European Physical Journal Web of Conferences*. Vol. 135. European Physical Journal Web of Conferences, p. 02001. DOI: 10.1051/epjconf/201713502001. arXiv: 1702.01395 [astro-ph.IM].

Monroe, Ryan (2018). *Github - Cosmic Ray Detection*. https://github.com/ryan-monroe/cosmicRayDetect.

Nelles, A., S. Buitink, A. Corstanje, E. Enriquez, et al. (2013). "Detecting radio emission from air showers with LOFAR". In: *American Institute of Physics Conference Series*. Ed. by R. Lahmann et al. Vol. 1535. American Institute of Physics Conference Series, pp. 105–110. DOI: 10.1063/1.4807530. arXiv: 1304.0976 [astro-ph.HE].

Nelles, A., S. Buitink, A. Corstanje, J. E. Enriquez, et al. (2015). "The radio emission pattern of air showers as measured with LOFAR - a tool for the reconstruction of the energy and the shower maximum". In: *JCAP* 5, 018, p. 018. DOI: 10.1088/1475-7516/2015/05/018. arXiv: 1411.7868 [astro-ph.HE].

Nelles, Anna (2018). Private Communication.

Offringa, A. R. et al. (2013). "The LOFAR radio environment". In: *Astronomy & Astrophysics* 549, A11, A11. DOI: 10.1051/0004-6361/201220293. arXiv: 1210.0393 [astro-ph.IM].

Parsons, A. R. et al. (2010). "The Precision Array for Probing the Epoch of Reionization: Eight Station Results". In: *The Astronomical Journal* 139, pp. 1468–1480. DOI: 10.1088/0004-6256/139/4/1468. arXiv: 0904.2334.

Pierre Auger Collaboration (2015). "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research A* 798, pp. 172–213. DOI: 10.1016/j.nima.2015.06.058.

Schellart, P., S. Buitink, et al. (2014). "Polarized radio emission from extensive air showers measured with LOFAR". In: *JCAP* 10, 014, p. 014. DOI: 10.1088/1475-7516/2014/10/014. arXiv: 1406.1355 [astro-ph.HE].

Schellart, P., A. Nelles, et al. (2013). "Detecting cosmic rays with the LOFAR radio telescope". In: *A&A* 560, A98, A98. DOI: 10.1051/0004-6361/201322683. arXiv: 1311.1399 [astro-ph.IM].

Schmidt, A. for the LOPES Collaboration (2009). "Self-Trigger for Radio Detection of UHCR". In: *Proceedings of the 31st ICRC, Lodz, Poland*. Vol. 31.

The Pierre Auger Collaboration (2016). "Nanosecond-level time synchronization of autonomous radio detector stations for extensive air showers". In: *Journal of Instrumentation* 11, P01018, P01018. DOI: 10.1088/1748-0221/11/01/P01018. arXiv: 1512.02216 [physics.ins-det].

Thoudam, S. et al. (2014). "LORA: A scintillator array for LOFAR to measure extensive air showers". In: *Nuclear Instruments and Methods in Physics Research A* 767, pp. 339–346. DOI: 10.1016/j.nima.2014.08.021. arXiv: 1408.4469 [physics.ins-det].

Tingay, S. J. et al. (2013). "The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies". In: *Publications of the Astron. Soc. of Australia* 30, e007, e007. DOI: 10.1017/pasa.2012.007. arXiv: 1206.6945 [astro-ph.IM].

Torr, P.H.S. and A. Zisserman (2000). "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry". In: *Computer Vision and Image Understanding* 78.1, pp. 138–156. DOI: 10.1006/cviu.1999.0832. URL: https://doi.org/10.1006%5C%2Fcviu.1999.0832.

Turin, G. (1960). "An introduction to matched filters". In: *IRE Transactions on Information Theory* 6.3, pp. 311–329. ISSN: 0096-1000. DOI: 10.1109/TIT.1960.1057571.

van Haarlem, M. P. et al. (2013). "LOFAR: The LOw-Frequency ARray". In: *Astronomy and Astrophysics* 556, A2, A2. DOI: 10.1051/0004-6361/201220873. arXiv: 1305.3550 [astro-ph.IM].

Zarb-Adami, K. et al. (2010). “Beamforming techniques for large-N aperture arrays”. In: *2010 IEEE International Symposium on Phased Array Systems and Technology*, pp. 883–890. DOI: 10.1109/ARRAY.2010.5613258.

Zatsepin, G. T. and V. A. Kuz’min (1966). “Upper Limit of the Spectrum of Cosmic Rays”. In: *Soviet Journal of Experimental and Theoretical Physics Letters* 4, p. 78.